-/ C 5170-1

Medical Care

October 1981
Vol. 19 No. 10
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Determining Health Insurance Coverage of Technology: Problems and Options

Diverging Trends in Hospitalization: Fact or Artifact?

Mode of Payment as a Predictor of Health Status, Use of Health Services and Preventive Health Behavior: A Report From the Los Angeles Health Survey

Use of Ambulatory Medical Care by the Poor: Another Look at Equity

Improving Hospital Discharge Data: Lessons From the National Hospital Discharge Survey

Hospital Medical Staff Organization and Quality of Care: Results for Myocardial Infarction and Appendectomy

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
The Rand Corporation

Gordon H. De Friese, Ph.D.
University of North Carolina

Brant E. Fries, Ph.D.
Yale University

Paul M. Gertman, M.D.
Boston University Medical Center

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Joel W. Kovner, Dr.P.H.
Kaiser Permanente Medical Care
Programs, Los Angeles

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

Steven A. Schroeder, M.D.
University of California,
San Francisco

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $40.00; special price U.S. students, $23.00; all other countries except Japan, one year, $50.00. Single copies $5.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1981 by J. B. Lippincott Company. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

October 1981
Volume 19, Number 10

CONTENTS

ORIGINAL ARTICLES

Determining Health Insurance Coverage of Technology: Problems and Options. *Barbara Greenberg and Robert A. Derzon* 967

Diverging Trends in Hospitalization: Fact or Artifact? *Harold S. Luft* 979

Mode of Payment as a Predictor of Health Status, Use of Health Services and Preventive Health Behavior: A Report From the Los Angeles Health Survey. *Alfred C. Marcus* 995

Use of Ambulatory Medical Care by the Poor: Another Look at Equity. *Joel C. Kleinman, Marsha Gold and Diane Makuc* 1011

Improving Hospital Discharge Data: Lessons From the National Hospital Discharge Survey. *Linda K. Demlo and Paul M. Campbell* 1030

Hospital Medical Staff Organization and Quality of Care: Results for Myocardial Infarction and Appendectomy. *Stephen M. Shortell and James P. LoGerfo* 1041

COMMUNICATIONS

Noninvestigational Use of Unapproved Drugs—Experience From the Scandinavian Countries. *Elina Hemminki* 1056

Screening for Noncompliance Among Patients With Hypertension: Is Self-Report the Best Available Measure? *Thomas S. Inui, William B. Carter and Roger E. Pecoraro* 1061

BOOK REVIEWS

Regulating Health Care: The Struggle for Control, edited by Arthur Levin. *Charles T. Heaney* 1065

Explorations in Quality Assessment and Monitoring. Volume I: The Definitions of Quality and Approaches to Its Assessment, by Avedis Donabedian. *James P. LoGerfo* 1066

Mental Health Services: What Happens to the Clients? by Alfred H. Schainblatt and Harry P. Hatry. *Gary L. Tischler* 1067

Nurse Practitioners: USA, by Harry A. Sultz, O. Marie Henry and Judith A. Sullivan. *Don E. Detmer* 1068

MEDICAL CARE
October 1981, Vol. XIX, No. 10

Diverging Trends in Hospitalization: Fact or Artifact?

HAROLD S. LUFT, PH.D.*

It has been noticed recently that hospital use by Blue Cross enrollees has fallen substantially in the last decade and that this decline is in marked contrast to relatively stable utilization reported by hospitals. The trends in length of stay in the two data sources are almost identical—all the differential is attributable to differing trends in admission rates: a sharp increase for all persons under 65 years of age in contrast to a moderate decline for Blue Cross members. These two trends are diverging at the rate of 2 per cent per year. Household survey data for the under-65 population with private health insurance also indicates an increasing admission trend, in contrast to the Blue Cross figures, even though Blue Cross covers half the insured population. About half of the differential trend can be accounted for by the increasing rate of duplicate coverage, which artificially increases the enrollments reported by carriers. The increased duplicate coverage is a reflection of more multiple-earner families, a shifting occupation mix, better fringe benefits, higher unemployment and smaller families. Part of the increased admissions for the total under-65 population is due to Medicare coverage for the disabled and to Medicaid. Most important, it is apparent that existing data provide an inadequate and potentially misleading picture of the most crucial measures of utilization.

ATTENTION has recently been drawn to the substantial decline in hospital use over the last decade for members of both Health Maintenance Organizations (HMOs) and Blue Cross plans.[1–3] The number of inpatient days for Blue Cross regular group members, that is excluding those with nongroup or Medicare complementary coverage, has fallen from 865 days per thousand in 1968 to 708 in 1978, a decline of 18 per cent.† This decline is in contrast to a rising trend in hospitalization for the overall population. Figure 1 presents the basic data concerning the rate of hospital use for Blue Cross regular group members in contrast to the rates for (1) the total popu-

---

* Associate Professor of Health Economics, Institute for Health Policy Studies, School of Medicine, University of California, San Francisco.

This research has been supported by grant 18-P-97556/9-01 from the Health Care Financing Administration.

Address for reprints: Harold S. Luft, Ph.D., Associate Professor of Health Economics, Institute for Health Policy Studies, School of Medicine, University of California, San Francisco, CA 94143.

† Two segments of the Blue Cross membership are excluded from these data: 1) persons with Medicare complementary coverage and 2) those with individually purchased policies. The regular group enrollment and utilization data exclude Washington State and Montana. Medicare complementary and individual enrollees are relatively high users of hospital care. More important, changes over time in their proportion of the Blue Cross enrollment would be reflected in the average utilization rate for all Blue Cross members.

0025-7079/81/1000/0979/$01.30 © J. B. Lippincott Co.

lation using nonfederal, short-term general and other special hospitals as reported by the American Hospital Association (AHA) and 2) the under-65 noninstitutionalized population using those hospitals, as reported by the National Hospital Discharge Survey (NHDS).[4–6]

There are important policy implications if this divergence is real, but this article will present evidence that much of the apparent reduction over time in hospital use by Blue Cross members is a statistical artifact. Moreover, the factors creating this artifact are likely to account for the similar downward trends in utilization among enrollees in other conventional insurance plans and in HMOs. Data artifacts may lead to incorrect analyses and subsequent policies. More important, some of the artifacts arise from real trends of substantial policy importance in themselves.

This study has certain aspects of a detective story with several culprits and no motives. The data in Figure 1 indicate clearly diverging trends in hospital use. The task at hand is to indentify plausible explanations for the divergence and examine the evidence to support, reject or partially accept each explanation. Unfortunately, most of the evidence is circumstantial and some readers are likely to be more convinced than others. Even I am not willing to consider the case closed and I welcome new evidence for alternative explanations.

The body of this article is divided into four sections. The first attempts to clarify those aspects of hospitalization that exhibit diverging trends. This involves decompos-

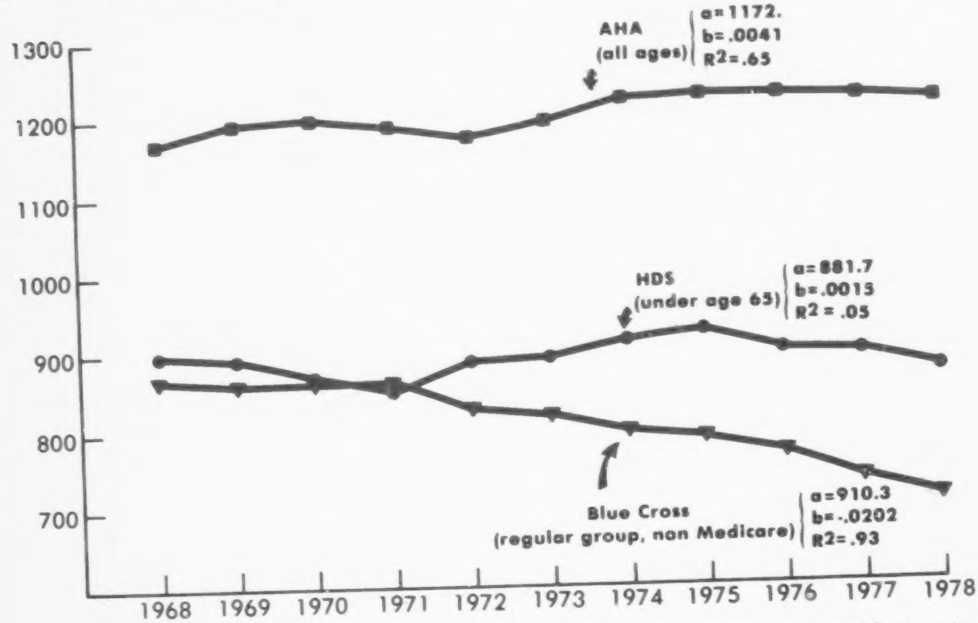


FIG. 1. Inpatient days per 1,000 persons. The AHA data are derived from annual statistics of the American Hospital Association and U.S. Census estimates of the total civilian population. National Hospital Discharge Survey (NHDS) data are from the NCHS survey of nonfederal short-stay hospitals, excluding newborns. Blue Cross data are for regular group members, excluding those with Medicare complementary coverage.

The standard formula for an annual rate of change is $y_t = a(1 + b)^t$, where y_t is the observed value in year t, a is a constant, b is the rate of change, and t is a time variable, equal to 1 in 1968, . . . 11 in 1978. The equation is linear in logs: $\log y_t = \log a + t \log(1 + b)$. The intercept, a, is the estimated value for 1967 (i.e., t = 0) and b represents the annual proportionate rate of change. For instance, b = −0.0202 means an annual rate of decrease of 2.02 per cent.

ing total hospital utilization rates into admission rates and length of stay, as well as excluding the over-65-population from the national data. Within this breakdown, it is important to determine whether hospitalization trends for Blue Cross enrollees are substantially different from the trends for people with commercial insurance coverage. The second section of the article examines the evidence of an artifact created by duplicate private insurance coverage that might explain part of the diverging trends in hospitalization. The third section introduces additional factors that may account for a portion of the remaining unexplained difference. A final section outlines some of the implications of the duplicate coverage problem for policy and analysis.

Clarifying the Sources of Divergence

The overall utilization data in Figure 1 can be split into trends in length of stay and admissions (or discharges) per 1,000.. As indicated in Figure 2, both BC and NHDS data show the average length of stay declining at about the same rate, 1.1–1.3 per cent per year. (This rate of change is derived from a regression line through the points for each year, and thus reduces the impact of year-to-year variability.)[‡][§] Figure 3, however, shows a very different pattern with a clear *increase* in discharges for the under-65 population as a whole (the NHDS data), in contrast to a more modest, but clear *reduction* in the Blue Cross data. The two trend lines for admissions are diverging by about 2 per cent per year, that is, the NHDS rate is increasing by 1.30 per cent per year while the BC rate is falling by

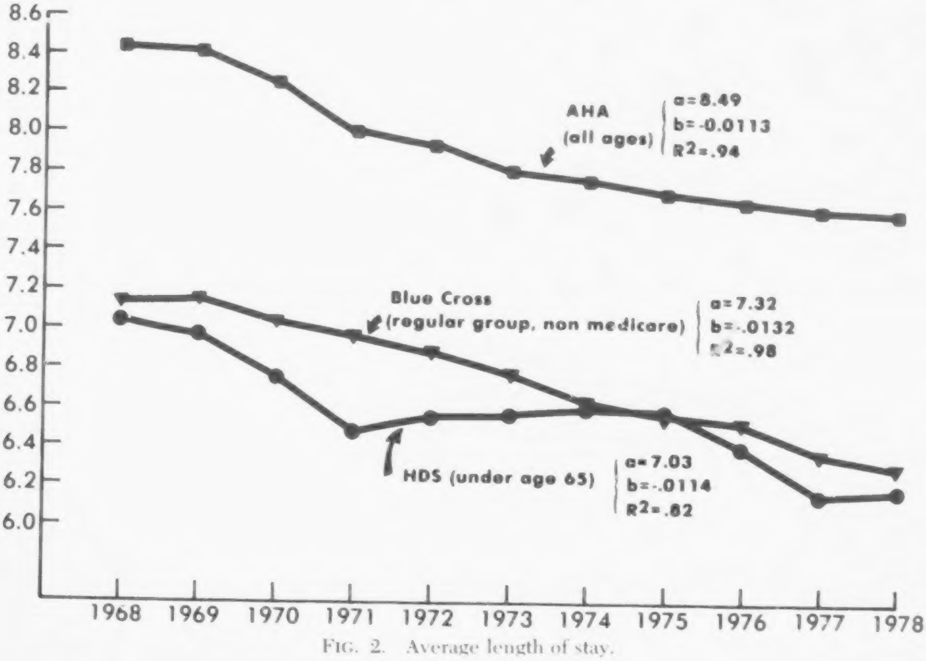


FIG. 2. Average length of stay.

[‡] See the notes to Figure 1.

[§] Some of the year-to-year variability in the NHDS is the result of its design. Since new hospitals are continually opening but the NHDS sampling frame is updated only every few years, the further one is from an updating year (1965, 1972 or 1975), the more the NHDS underestimates the true discharge rate.[7]

0.67 per cent. The strong similarity in both the levels and trends, for length-of-stay supports the notion that the two data sets are capturing similar types of and changes in hospital use. Since the LOS calculation depends only on patient days and admissions, it is not subject to errors in the estimation of enrollments. This suggests that the impressive reduction in stay over the past decade is not only real, but has been experienced by both Blue Cross and non–Blue Cross members.[‖] It is possible that Blue Cross enrollees experienced an even greater reduction in case-specific length of stay that was offset by changes in case mix. However, there is no evidence to support or reject this possibility.

An initial criticism of this analysis is that the NHDS includes all discharges from short-term non-Federal hospitals, regardless of coverage. By limiting the analysis to the population under age 65, the confounding influence of most Medicare beneficiaries can be eliminated, although under-65 Medicare disabled beneficiaries are included. Blue Cross accounts for about a third of all discharges among the nonelderly population. Therefore, if all the apparent BC decline were true, the average annual increase in admissions for the non-BC segment of the under-65 population would be 2.27 per cent, in contrast to the 1.30 per cent seen from the NHDS (Fig. 3).[#] If this increase is attributed to a subset of the non-BC population, the rate of divergence becomes even greater. For such a rapid increase in hospital admissions among the non-BC commercially insured

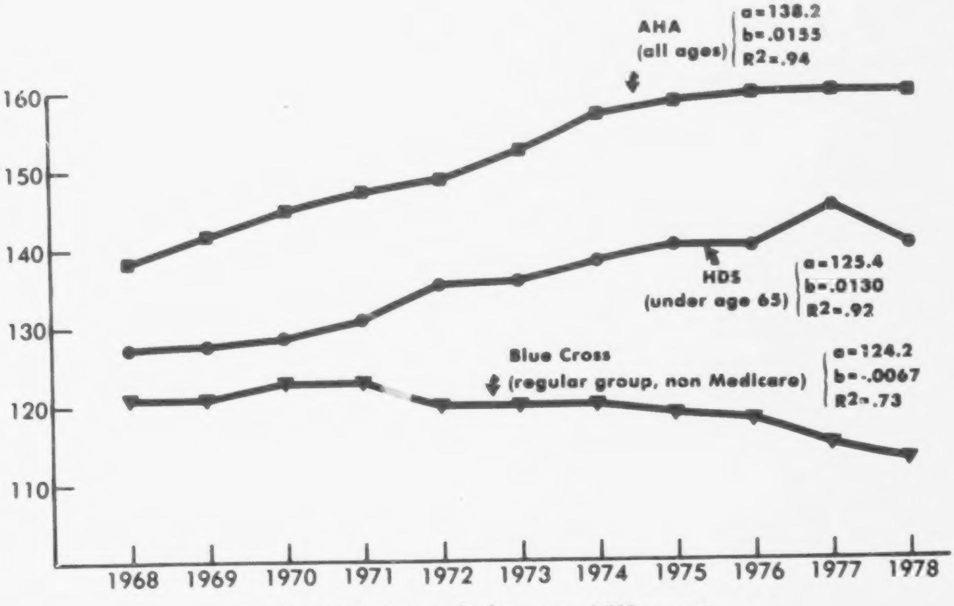


FIG. 3. Admissions or discharges per 1,000 persons.

[‖] For instance, during the period 1968 to 1977 the average length of stay for enrollees in the two California Kaiser plans fell at the rate of 1.56 per cent per year. This somewhat more rapid decline may be attributable to several factors, including the relatively greater importance of maternity admissions in HMO populations.

[#] This calculation is as follows: the one-third of the under-65 NHDS total represented by BC has an annual decline of 0.67 per cent while the total under-65 NHDS is increasing by 1.30 per cent. Therefore, letting x = annual change for the non-BC population,

$$\begin{aligned} 0.33(-0.67) + 0.67x &= 1.30 \\ 0.67x &= 1.52 \\ x &= 2.27. \end{aligned}$$

population to have gone unnoticed and unreported would be surprising, but not impossible. The limited availability of data suggests that it may be difficult to monitor such changes in true hospital-use rates.

Although hospital utilization data for people with commerical coverage are not available, there is an indirect test of the possibility that a falling Blue Cross hospitalization rate is being offset by a rising commercially insured rate to produce the observed NHDS results. If this were true, it would probably be reflected in claims data. (The major reason to expect different patterns for Blue Cross and commercial enrollees is a "catching-up" in improved coverage among the latter.) In fact, for 1968-73, the period during which comparable data are available, hospital benefits per Blue Cross enrollee rose at an annual rate of 11.4 per cent in contrast to 8.2 per cent for commercially covered people.[8,9] To the extent that the charges paid by commercial carriers rose more rapidly than costs paid by Blue Cross, these rates underestimate the discrepancy. In any case, there is no evidence to support the notion of offsetting trends between commercial and Blue Cross enrollees. However, a subsequent section of this article will provide evidence of important increases in hospital use by small segments of the population *without* private insurance coverage.

While the NHDS data do not separate people with private insurance, there is some fragmentary evidence for the under-65 population from another source. The National Health Interview Survey (NHIS) occasionally provides population-based estimates of hospitalization rates for the insured population. These NHIS data are household-based and thus are not subject to double counting.¶ In 1968, 9.01 per cent of all persons with private hospital insurance aged 64 and under reported a hospitalization; this led to a total of 733 days per 1,000 persons per year.[10] In 1974 the comparable figures were 9.90 per cent and 851 days per 1,000.[11] These data yield an increase similar to that of the NHDS, and the rising use is inconsistent with the BC data, even though BC members represent more than half of the privately insured population.[11]

The Duplicate Coverage Problem

The computation of a hospital admission rate obviously depends upon numerators and denominators. The Blue Cross admission rates represent covered inpatient stays divided by Blue Cross enrollment.[12] The AHA and NHDS rates are computed from discharges reported by hospitals and the relevant civilian population (total, or under-65) estimated by the U.S. Bureau of the Census.[4-6] Although one might expect the Blue Cross rates to be more accurate because numerator and denominator are derived from the same organization, consideration of other factors suggests this is not the case. Enrollment figures for Blue Cross, or for any single insuring organization, are not adjusted for the increasing extent of duplicate health insurance coverage in the United States. While carriers may attempt to coordinate benefits when paying claims, they do not count people with multiple policies as fractional enrollees. Therefore, the enrollment figures used in the denominators are overstated. Moreover, duplicate coverage has been rising in the last decade, primarily because an increasing proportion of married women are employed in the paid labor force and because more of these women are working in jobs with employer-sponsored group coverage. A second important factor is that enrollment figures have not adequately tracked the declining size of the American family, again overstating the denominator.

¶ There are some problems with NHIS data because of nonresponse, forgetfulness, and, most important, the exclusion of data for people dying before the survey is conducted. While these problems will influence the magnitude of the hospitalization rate, there is no reason to expect these biases to affect the trend analyses reported here.

The Magnitude and Growth of Duplicate Coverage

In attempting to clarify the hypothesized problem with the denominator, the situation becomes more complex because there is no simple, accurate way to determine how much duplication truly exists. One method is to use the National Health Interview Survey to ask people whether they have more than one policy for hospital coverage. The proportion responding affirmatively rose from 11.4 per cent in 1968 to 13.0 per cent in 1974, with 0.8 per cent of the latter reporting three or more plans.[10,11] While this indicates that duplicate coverage is real and growing, it probably understates the actual increase because of biases inherent in household surveys. Survey estimates of coverage may be biased downwards because people may not know or may forget that they have coverage.[13–15] It seems even more likely that people will forget *how many* plans provide coverage. This suggests that, as duplicate coverage becomes more common, it may be increasingly underestimated by surveys unless the survey is carefully designed to track down all types of coverage.[16]

Alternatives to survey estimates are provided by the Health Insurance Institute (HII).[8] The HII figures are derived from reports by carriers of the number of people covered under various policies, with adjustments to remove what HII estimates is duplicate coverage. The HII coverage figures are biased upwards because this adjustment is probably incomplete and because of the lag between the time someone leaves a job and loses coverage, and the time the insurer adjusts its records. This results in phantom enrollments—people whose ineligibility is discovered only if a claim is submitted.**

Although the choice between the HII net coverage and survey estimates is important for determining the number of people without coverage, the estimate of *changes* in the level of duplicate coverage is rather insensitive to the choice of data series. Table 1 presents various estimates of the trends in gross and net hospital insurance coverage. The periodic surveys of the National Center for Health Statistics provide unduplicated counts of the number of people with private coverage. There has been some increase in the percentage of the population with coverage—it rose from 75.9 per cent in 1968 to 76.8 per cent in 1976. This increase of less than 1 per cent is small in contrast to year-to-year variations of greater magnitude reflecting either changes in the questionnaire (1972) or substantial increases in unemployment (1975). To minimize the influence of such year-to-year variations, annual rates of increase are computed for each of the data series.

The gross enrollment figures published by the Health Insurance Institute are more problematic because of substantial changes in reporting. The annual HII data include the number of people with four different sources of hospital insurance: 1) individually sold policies written by insurance companies; 2) group policies from insurance companies; 3) Blue Cross–Blue Shield; and 4) independent and other plans, largely employer and union programs and HMOs. HII attempts to remove duplicate coverage *within* each source, but it is unclear how this is done. For instance, the HII estimate for gross enrollment in 1973 was nearly 30 million lower than for 1972, reflecting an adjustment in the esti-

---

** The HII undertakes periodic samples of claims to determine the proportion of people who say they have duplicate coverage, and thus are eligible for coordination of benefits. Since there are obvious reasons for not reporting duplicate coverage, one can assume that such surveys underestimate duplication, much like tax returns underestimate incomes. The HII adjusts upwards these duplication estimates to account for the underreporting, but these adjustments are not based on hard figures. The fact that the HII periodically reduces retroactively its estimates of the net number of people with coverage supports the notion that their coverage figures have an upward bias.

mate for individual coverage. When the HII makes such a change, estimates of coverage for the total population in earlier years are recomputed, but this retroactive adjustment is not made for the population under age 65. Thus, the series for that population group exhibits discontinuities, particularly when individual policies are included (column 5).

The duplication rate is defined as gross enrollment divided by net enrollment. Thus, a rate of 1.25 means that enrollments reported by carriers are 25 per cent above the number of people with coverage reported in household surveys. This implies that their admission rates will have denominators 25 per cent larger than they should be. The annual rate of increase in the duplication rate is equal to the difference in the annual rates of increase between the numerator and denominator—that is, the HII and NCHS figures, respectively.††

Because there are several sets of figures, there is potential for controversy in the choice of data.‡‡ Fortunately, the results using various reasonable alternatives are all quite close. Ideally, the duplication rate would be measured just for people with regular group Blue Cross coverage. (Note that this would be in the denominator; the numerator would include all the policies held by those people.) Such data do not exist. The closest approximation is the NCHS series of net coverage for the under-65 population, column 2 in Table 1. The rate of increase in such coverage, 0.68 per cent per year, may be compared with the increase in gross enrollments for the under-65 population, columns 5 and 6. Enrollment under "all but nongroup commercial policies" for these people (column 6) has been increasing at 1.70 per cent per year, or 1 per cent more rapidly.§§ (Column 6 – Column 2) As noted, a consistent set of data is not available for all policies held by the under-65 population (column 5), so an annual rate of change cannot be meaningfully computed. The two HII series for all age groups, columns 3 and 4, exhibit remarkably similar rates of increase: 2.08 per cent and 2.21 per cent. Moreover, the annual increase in duplicate coverage for all persons (column 3 or 4 minus column 1) is also about 1 per cent per year, regardless of the HII series used.

Thus, the best estimates of the annual rates of increase in reported enrollments exceed true coverage increases by about 1 per cent per year. The enrollment figures are crucial to the calculation of admission rates; therefore this duplicate coverage problem accounts for half the observed 2 per cent difference between admission-rate trends reported by Blue Cross and the National Hospital Discharge Survey.

The preceding analysis rests on the crucial assumption that the growth rate in dup-

---

†† The derivation of this is as follows: let G and N be the gross and net enrollment levels at a point in time, so G/N is the duplication rate. The rate of increase in G/N is $\frac{d\,(G/N)}{dt} = \frac{dG}{dt} - \frac{dN}{dt}$. Since G and N can be expressed as $G = a_0e^{b_0t}$ and $N = a_1e^{b_1t}$, $\frac{dG}{dt} = b_0$, $\frac{dN}{dt} = b_1$. Thus, $\frac{d\,(G/N)}{dt} = b_0 - b_1$.

‡‡ In fact, the numbers *within* each column are open to question. The estimates from NCHS of the unduplicated persons with coverage (column 2) are subject to random sampling error and an adjustment for unknown coverage status. (The latter is a proportionate allocation based on people with known coverage.) Similarly, the HII figures are only as good as their underlying sources. Some Blue Cross plans take special efforts to discourage duplicate BC coverage while others purposely double-count people with known duplicate coverage.[17,18] While problems undoubtedly exist, one cannot determine their likely impact.

§§ This category is the sum of the following three types of coverage reported by the HII: 1) group policies written by commercial insurance companies, including Administrative Service Only agreements and Minimum Premium Plans; 2) Blue Cross–Blue Shield and Medical Society Plans; and 3) Other Plans. Excluded are individual and family policies written on a nongroup basis by commercial carriers.

TABLE 1. Gross and Net Enrollment in Private Health Insurance Plans
1968–77 (000 omitted)

	Net Enrollment—Unduplicated Number Persons with Coverage		Gross Enrollment			
			All Ages		Under Age 65	
	All Ages (1)	Under 65 (2)	All Policies* (3)	All But Nongroup Commercial Policies* (4)	All Policies† (5)	All But Nongroup Commercial Policies† (6)
Year						
1968	150,888‡	141,572‡§	176,647	151,369	183,626	147,175
1969	NA	NA	184,289	158,367	191,655	154,034
1970	154,063‡	143,611‡‖	190,349	163,691	198,590	158,995
1971	NA	NA	193,498	165,725	203,156	160,567
1972	155,253‡ #	143,309‡ #	196,796	167,785	205,720	161,634
1973	NA	NA	200,710	173,166	186,285	163,568
1974	161,104#	148,349#	206,646	177,839	192,093	168,725
1975	162,378¶	147,295**	212,828	182,713	196,668	172,428
1976	164,235¶	149,643¶	209,648	182,652	193,247	171,510
1977	NA	NA	213,930	185,243	195,013	173,384
Time trend regressions††						
a = Intercept	148,821	140,281	176,933	151,183		148,506
b = Annual rate of increase	0.0109	0.0068	0.0208	0.0221		0.0170
R^2	0.9707	0.8834	0.9533	0.9684		0.9340

NA: not applicable.

[*] Health Insurance Institute, Source book of health insurance data, 1978–79. Washington, D.C.: Health Insurance Institute, 1979.

[†] Health Insurance Institute, Source book of health insurance data. Washington, D.C.: Health Insurance Institute, yearly, 1968–79. For the years 1968–72 the enrollment figures for persons under age 65 with all types of policies (column 5) exceed the corresponding figures for all persons (column 3) because the latter figures have been retroactively adjusted to reflect new HII estimates of nongroup coverage. This retroactive adjustment is not available for the population under age 65. Each volume of the Source Book includes data for both the preceding year and revised figures for the year before that. With the exception of 1977, these data are the revised figures.

licate coverage for Blue Cross is equivalent to that for all private health insurance. (The absolute level of duplication could be higher or lower; only the rate of change needs to be the same.) Direct tests of this assumption are not available, but two bits of evidence provide some support. For 1968-77 the growth rate in enrollment in Blue Cross plans is 2.41 per cent per year, in contrast to 1.85 per cent for commercial insurers.[8,9,‖] If duplicate coverage were growing *more* rapidly among commercial carriers, their gross enrollment would probably also be growing more rapidly than that of Blue Cross. The second piece of evidence is drawn from an analysis of cross-state variations in duplicate coverage.[19] After controlling for several factors that explain duplicate coverage, such as occupational mix, Blue Cross market share within the state had no impact on the ratio of gross to net coverage in the state. If Blue Cross plans were more effective in eliminating duplicate coverage, a negative coefficient would be apparent in the regression. In fact, the coefficient was positive and insignificant.

Explanations for the Increasing Rate of Duplicate Coverage

Several factors influence the rate of duplicate coverage for people with group policies.[¶¶] In many married couples both husband and wife may cover each other. Such duplication may be intentional— to improve coverage through double claims or coordination of benefits. But duplication may also be involuntary. More than two

---

[‖] The annual growth rate for commercial insurers, 1.85 per cent, refers to group policies only, so as to avoid the problems with individual policies. The growth rate for individual plus group coverage was only 1.71 per cent.

[¶¶] People may supplement group policies with individual coverage. However, the group policy will usually provide primary benefits for hospitalization, so such duplication will have a small influence on estimated utilization rates under group policies.

[‡] Mueller MS. Private health insurance in 1973: a review of coverage, enrollment and financial experience. Soc Sec Bull 1975;38(2):21–40.

[§] U.S. National Center for Health Statistics. Hospital and surgical insurance coverage: United States—1968. Vital and health statistics. Series 10, number 66. Washington, D.C.: U.S. Government Printing Office, 1972. (DHEW publication no. (HSM) 72-1033.

[‖] U.S. National Center for Health Statistics. Monthly vital statistics 21:9 (supplement), December 18, 1972.

[#] U.S. National Center for Health Statistics. Hospital and surgical insurance coverage, United States—1974. Vital and health statistics. Series 10, number 117. Washington, D.C.: U.S. Government Printing Office, 1977. (DHEW publication no. (HRA) 77-1545.) It should be noted that the insurance question for 1972 was worded differently from previous years and was more difficult to process. Therefore, the 1972 estimates are probably biased downwards. See page 4 in this reference.

[¶] Carroll MS. Private health insurance plans in 1976: an evaluation. Social Sec Bull 1978;41(9):3–16.

[**] U.S. National Center for Health Services Research and National Center for Health Statistics. Health: United States—1978. Washington, D.C.: U.S. Government Printing Office, 1978:409. (DHEW publication no. (PHS) 78-1232.)

[††] These regressions are based on the equation $y = a \cdot (1 + b)^t$, where y = annual enrollment, b = annual rate of increase, and t is a time variable, with 1968 = 1; 1977 = 10. The coefficients are estimated after a log transformation, $\log y = \log a + t \log (1 + b)$.

thirds of the workers in private industry with hospital benefits have them paid exclusively by employer contribution.[20] Since the employee usually does not have the option of transferring unneeded duplicate health insurance premiums into other fringe benefits, there is no incentive to decline such coverage. In other instances, duplication may provide valuable complementarity, such as belonging to both a Health Maintenance Organization for first-dollar coverage of ambulatory services and to a conventional insurance plan to maintain coverage for second opinions. If coverage is heavily subsidized such duplication may be worth a small additional premium.

Duplicate group coverage also occurs in ways other than the multi-earner family. A substantial number of people (nearly 4 million in 1976) hold multiple jobs.[5] Although most employer-sponsored health insurance is available only to employees working half-time or more, some multiple job holders are likely to be eligible for such coverage. Others, such as construction workers, must accrue a specified number of hours per time period, such as 320–399 hours per 6 months, to maintain coverage.[21] With such relatively low coverage criteria, it may be possible to combine part-time employment with a full-time job elsewhere.

An administrative form of duplicate coverage occurs through job turnover. Many employees retain coverage for a specified period after leaving an employer. In some cases this eligibility may overlap the coverage obtained from a new employer. (However, some employers have waiting periods for new employees.) Even if the grace period has expired, the carrier may not check the employee's eligibility until the next contract period unless a claim is filed. Thus, there are a number of phantom enrollees through reporting lags.

Social and economic changes in the last decade have altered these factors in ways that increase the duplicate coverage rate. The proportion of married women in the labor force has been steadily rising. It was 47.6 per cent in 1970 and 54.8 per cent in 1976, an annual increase of nearly 2.5 per cent.[5] Among full-time employed workers in 1972, 76 per cent of all married men and 57 per cent of all married women had group coverage through their employers. More important, of those with group coverage, 90 per cent of the married men covered their dependents, as did 63 per cent of the married women.[22] Women have also been changing their occupations—moving from those with relatively poor health insurance coverage to those with better coverage. For instance, using the coverage rates in 1972 for each occupation, the shift in women's job mix from 1965 to 1976 would have increased their group coverage rate from 56 to 59 per cent.[22]

Throughout the 1960s and 1970s employers have been increasing their contributions to fringe benefits more rapidly than wages. Part of this increase has taken the form of eliminating the employee contribution for health insurance.[23] Another aspect is the increasing number of employer plans that provide complete coverage for dependents without additional employee contribution. Both trends will increase the duplication rate.

Other administrative factors may also explain part of the increase in duplicate coverage. The unemployment rate among married men had markedly increased, from 2.1 per cent in 1967-68 to 5–6 per cent in 1975-76.[24] This suggests more phantom enrollees because of job changing. The problem of inaccurate reporting may be exacerbated by the rapid rise in Administrative Services Only (ASO) and "self-certification" arrangements whereby firms self-insure and use carriers primarily to handle claims and benefits paperwork.[8] Carriers involved in such arrangement have even less incentive to keep careful track of enrollment changes and in some instances do not keep enrollment files, instead forwarding claims to the employer to determine eligibility.

A final aspect of the duplicate coverage problem arises from the fact that few carriers have an actual head count of their enrollees, irrespective of their enrollees' other coverages. Instead, carriers know how many contracts are in force, and whether they cover one person or a family. (Some family contracts are further classified by two-person and three-or-more-person coverages.) Occasionally, surveys may be undertaken to determine how many people are covered under the average family contract, but such surveys depend upon the carrier and are infrequent. The ratio of persons covered to primary insureds (i.e., contracts) was constant at 2.459 from 1965 to 1970 and fell slightly to 2.423 by 1975.[8] In contrast, average U.S. family size fell from 3.70 in 1965 to 3.39 in 1976.[5] More important, the ratio of persons aged 0–64 to workers aged 16–64 fell from 2.374 in 1965 to 2.017 in 1975, a 15 per cent decline, in contrast to the 1.6 per cent decline in coverage ratio reported by the carriers.[25] This lag in adjusting the family factor leads to an overestimate of the number of persons having insurance.

There have also been marked changes in family structure and enrollment structure that are hidden by the aggregate statistics. The number of households headed by women with children has increased 62.3 per cent from 1970 to 1977.[27] Since children are low utilizers of hospital care, this change in enrollee mix may have an impact on measured use. For some plans, the proportion of single contracts had been rising, which implies that the relatively constant ratio of enrollees/contract reported by the HII is even more biased for family coverage.

Additional Factors Potentially Explaining the Diverging Trends in Hospital Admission Rates

The estimated duplicate coverage effect accounts for only half the differential in hospital admission rates. Part of this shortfall may be attributed to the fact that the difference in utilization trends is based on Blue Cross figures while the duplicate coverage rates refer to all insurers. If commercial carriers had more accurate counts of their enrollees, either through more realistic family factors or fewer phantom enrollees arising from job changes, then the true duplicate coverage rate for Blue Cross enrollees would be higher than for all enrollees, and thus would explain a larger fraction of the utilization difference. Employees with Blue Cross coverage tend to have less restrictive eligibility requirements than those with commercial coverage and this may extend to post-employment coverage as well.[21] On the other hand, if Blue Cross plans had more accurate counts of their enrollees, then even more of the differential is unexplained.*

Blue Cross plans tend to offer more extensive coverage[20] but commercial carriers have been expanding their benefits over the 10-year period and have increased the proportion of actual days paid by the plan.[8] This benefit expansion could have resulted in increased utilization and, more important, such an increase might be expected and thus not a source of concern. But this explanation does not fit the more rapid growth in hospital benefits per enrollee experienced by Blue Cross enrollees. At the same time, improved coordination of benefits among plans may have reduced the number of hospital days paid for twice. This would reduce the BC numerator, although not the actual number of days reported by the AHA or the NHDS.

There is also evidence of increased hospital utilization by people not covered by private health insurance.† Since the utilization differentials appear for the population under 65 years of age, we need not be concerned with Medicare use by the elderly. However, effective in 1973 Medicare coverage was extended to disabled beneficiaries receiving Social Security cash benefits. This could have had a twofold effect. First, the disabled tend to have a high need for medical care, yet in 1972 only 38 per cent had private hospital coverage.[27] The new Medicare coverage may have resulted in an increased use of hospital services, increasing the AHA/HDS figures without altering the BC patterns. Second, Medicare coverage is always treated as primary, so even if the disabled retained their Blue Cross policies, the hospitalization would be attributed to Medicare. While the utilization rate for such people may not have increased substantially, the effective transfer of such high users out of Blue Cross would lower the average BC rate.

---

* However, as the previously discussed evidence from enrollment statistics suggests, the Blue Cross duplicate coverage rate was probably growing more, rather than less, rapidly than that for commercial carriers.

The number of hospital days for Medicare disabled beneficiaries rose from 2.5 million in the second half of 1973 to 9.6 million in all of 1977. Discharges rose from 233,020 in 1973 to 937,245 in 1977.[28] The use by Medicare disabled beneficiaries represents about 10 per cent of all patient days attributable to the under-65 population.[29] People covered under the End Stage Renal Disease Program account for about 6 per cent of the discharges of these Medicare disabled beneficiaries. While this is substantial, we cannot assume that none of these patients received treatment prior to Medicare coverage, nor that all switched from Blue Cross coverage. Some must have been covered by commercial insurers and many had no coverage at all. Moreover, the divergence in BC and NHDS utilization is evident prior to 1973 (Fig. 3.)

The second major source of public funding is the Medicaid program. Unfortunately, national statistics on hospital admissions and days are not available. Table 2 presents an estimate of Medicaid admission rates derived from expenditures on inpatient services in general hospitals and the average expense per admission. Although these are only rough estimates, they are of the right order of magnitude. The figures indicate a moderate increase of 2.3 admissions per 1,000 population between 1968 and 1972, with a small decline thereafter. The comparable increase for the same period in the NHDS was 8.3 admissions per 1,000 population.

A third potential explanation for the differential is the increasing use of general hospitals for psychiatric care. Among the population under age 65, the number of discharges with diagnosis of mental disorders per 1,000 population rose from 5.38 in 1971 to 7.54 in 1977.[6] Almost all group insurance plans provide coverage for confinement in general hospitals for mental disorders, so a major portion of this increase is probably covered by Blue Cross and thus cannot account for the differential.[20] The increasing role of federal- and state-funded community mental health centers may have had an influence on the increased use by the noninsured population.[30]

What does all this mean with respect to interpreting the different hospitalization trends reported by Blue Cross and the NHDS? While Blue Cross incentives and controls may have had a substantial differential impact on hospital use, it seems premature to attribute all the divergence in trends to such policies.‡ (Note that to ex-

---

† More precisely, there was increased use by people with public coverage through Medicare and Medicaid. Such people may or may not have private coverage. Utilization of federal hospitals, such as Veterans Administration facilities, is excluded from both the NHDS and BC figures so changes in such use should not alter the analysis.

‡ Moreover, local cost-control innovations that may be supported by logic and initial evaluations sometimes have little or no overall effect. For instance, recent evaluations suggest that a second surgical opinion program has actually increased surgery rates.[31]

TABLE 2. Estimated Use of General Hospitals by Medicaid Beneficiaries—1968–1977

Fiscal Year	Medicaid Inpatient Services in General Hospitals*† (in Millions)	Average Expense Per Adjusted‡ Admission	Estimated Medicaid Admissions (Thousands)	Estimated Medicaid Admissions Per 1,000 Persons Under Age 65§
1968	$1,361	$471.30	2,888	16.3
1969	1,586	539.25	2,941	16.5
1970	1,887	610.10	3,093	17.1
1971	2,288	675.01	3,390	18.5
1972	2,557	744.88	3,433	18.6
1973	2,660	796.65	3,339	18.0
1974	2,887	878.84	3,285	17.6
1975	3,411	1016.79	3,355	17.9
1976	3,938	1168.15	3,371	17.9
1977†‖	3,678	1316.05	2,794	14.7

* U.S. House of Representatives. Committee on interstate and foreign commerce. Data on the Medicaid programs, eligibility, services, expenditures, fiscal years 1966–77. Washington, D.C.: U.S. Government Printing Office, 1977:32 (years 1968-71).

† U.S. Health Care Financing Administration. HCFA program statistics. Health care financing review 1979:1(2):74. 1972–76 are final figures; 1977 is preliminary.

‡ U.S. NCHS and NCHSR. Health: United States—1978. Washington, D.C.: U.S. Government Printing Office, 1978:416. (DHEW publication no. (PHS) 78-1232.)

§ Civilian noninstitutionalized population under age 65 from U.S. NCHS. Current estimates, yearly.

‖ Expenditures adjusted for the 15-month fiscal year ending September 1977.

plain the differences in trends, the policies must have had a differential impact on BC patients. For instance, for the same medical problem, a physician would have to be more reluctant to admit a patient with Blue Cross coverage than one with, say, Aetna or Prudential coverage.) Instead, about half the difference in trends may be explained by the increasing rate of duplicate coverage, and thus reflect a data artifact. Other parts of the difference may be explained by increasing public coverage through Medicare and Medicaid and publicly funded community mental health centers. While people with public coverage are a relatively small fraction of the population, they represent a disproportionate share of hospital use. Thus, it is not really appropriate to compare the use of people with private coverage to the totals in the HDS and AHA data. Once all these other nonquantifiable factors are considered, it is difficult to determine the true magnitude of the Blue Cross policy changes.

Implications for Policy and Analysis

The increasing importance of duplicate health insurance coverage has implications for both policy and analysis. The policy implications relate to rate setting and to the design of certain approaches to financing health insurance. The analytic problems arise for both academic researchers and in-house analysts monitoring plan performance.

Prospectively determined insurance rates are usually based upon current utilization rates and projected costs. To the extent that utilization per enrollee falls because of duplicate coverage, projected premiums will exceed actual costs. Although the utilization trend for BC members is clearly downward sloping, an actuary would conservatively use the most recent estimate rather than assume a continuation of the trend. If all premiums are effectively experience-rated, the surplus will be returned. One may question, however, whether true cost-containment measures are more likely to be undertaken in situations of artificial surpluses or in situations of real budget constraints. Surpluses may increase the ability to finance new adminstrative controls, but the tightness of such controls probably reflects budget pressures.

Even if all premiums are experience-rated, the duplicate coverage problem may lead to inequities based upon factors other than utilization differences. The cost of health insurance per enrollee will be lower among employees in firms with a high proportion of working spouses. (In some ways this is analogous to the "marriage tax" incorporated in the federal tax laws.) More important, the second earner in a family gets relatively little from his or her employer contribution for health insurance—that fringe benefit actually accrues to others through lower premiums. This situation would be altered radically by allowing workers to shift the employer contribution to other types of fringe benefits. For instance, if one spouse chose family health insurance coverage, the other might use the comparable employer contribution for a retirement plan. Obviously, such a shift might increase costs to the employer. Under the present system, a firm with a large proportion of married workers might save a substantial amount by tailoring health benefits so that married people would choose to be covered under their spouse's plan, thereby lowering the experience-rated costs to the firm. Any scheme for financing national health insurance through employer contributions must address the situation of multiearner families.

The findings of this article also raise serious problems for analysts. The experienced researcher knows that data are not always what they seem to be at first glance. Yet, finding problems with such crucial figures as hospitalization rates should give one pause. For instance, much of the discussion of the competitive impacts of HMOs has analyzed declines in hospital use by Blue Cross enrollees or lower hospital use by Blue Cross enrollees in areas with high HMO penetration.[32–34] Yet duplicate coverage may be directly related to HMO enrollment for two reasons. Historically, HMOs have developed in areas with extensive employer contributions, making duplicate coverage less expensive to the enrollee. In some cases, HMO and conventional coverage tend to complement one another, making such duplicate coverage more desirable than duplicate conventional coverage.§

Problems also arise for the monitoring of cost-containment activities of insurers. For instance, it becomes much more difficult for a plan to evaluate a preadmission screening program, ambulatory surgical coverage, or outpatient benefits if there are concurrent, unmeasured changes in the local duplicate coverage rate.[3] Since duplicate coverage is highly dependent on local occupational and labor market factors, one cannot easily apply national trends or use "matched" controls.[19]

---

§ To give an extreme example, the health insurance market in Hawaii is dominated by Kaiser and Hawaii Medical Services Association (HMSA). HMSA coverage typically includes some copayments and if someone has multiple HMSA coverages, only the most comprehensive plan applies; coordination of benefits is not used to provide 100 per cent coverage. Hawaii has an extremely high rate of multiearner families, many of whom have dual choice (Kaiser or HMSA) options. Since multiple HMSA coverage provides no additional benefits, one might anticipate a very high rate of duplicate coverage. This may explain the rapidly falling hospitalization rates for both Kaiser and HMSA in spite of an increasing hospitalization rate reported by the AHA. The decision to have HMO plus conventional coverage is actually very complex. It depends on the types of plans involved, the family's usual sources of care and the net premium costs. While a clear understanding of these enrollment factors is not yet available, the potential for duplicate coverage must be considered.

Having recognized the problem, is it possible to do something about it? If policy changes are undertaken to encourage the flexible use of fringe benefits for things other than health insurance, the problem may disappear. The identification of a statistical artifact is unlikely to have such a policy impact, but the recognition that duplicate coverage is a large and growing factor might help efforts in that direction. While the problem remains with us, the National Center for Health Statistics might attempt to measure more accurately the extent of multiple coverage. At a minimum, insurers could attempt to maintain more accurate head counts of the number of eligible enrollees. Some carriers already require the use of a unique identifier, such as the Social Security number. Alternatives could be designed to protect individual privacy while allowing computerized matching of people covered by multiple carriers. If carriers know how many people have multiple coverage, they can then adjust the enrollment statistics accordingly. Until such changes are made, however, users of such data should interpret utilization rates with extreme caution.

Acknowledgment

I am indebted to William Elliot, Susan C. Maerki, Wendell Primus, Steven Schroeder, Keith Winnard and anonymous reviewers from *Medical Care* and *Inquiry* for their valuable comments.

References

1. Blue Cross Association. Financial outlook 8. October 1977.

2. Luft HS. Trends in medical care costs: Do HMOs lower the rate of growth? Med Care 1980;18:1.

3. McNerney W. Blue Cross plans experience sharp 10-year decline in hospital utilization rates. Chicago: Blue Cross and Blue Shield Associations, January 18, 1980.

4. American Hospital Association. Hospital statistics. 1979 ed. Chicago: American Hospital Association, 1979.

5. U.S. Bureau of the Census. Statistical abstract of the United States 1978. Washington, D.C.: U.S. Government Printing Office, 1978.

6. U.S. National Center for Health Statistics. Utilization of short-stay hospitals, summary of nonmedical statistics. Vital and health statistics, series 13, 1968–77.

7. Moien M. The use of hospital discharge survey data to produce trend estimates of hospital utilization. American Statistical Association 1978. Proceedings of the Social Statistics Section. Washington, D.C.: American Statistical Association, 1979:210–15.

8. Health Insurance Institute. Source book of health insurance data. Washington, D.C.: Health Insurance Institute, annual.

9. Blue Cross and Blue Shield Associations. Fact Book. Chicago: Blue Cross and Blue Shield Associations.

10. U.S. National Center for Health Statistics. Hospital and surgical insurance coverage, United States, 1968. Vital and health statistics, series 10, no. 66. (DHEW publication no. (HRA) 72-1033, 1972.)

11. U.S. National Center for Health Statistics. Hospital and surgical insurance coverage, United States, 1974. Vital and health statistics, series 10, no. 117. (DHEW publication no. (HRA) 77-1545, 1977.)

12. Blue Cross Association, Health Economics Section. Personal communication.

13. Mueller MS. Private health insurance in 1973: a review of coverage, enrollment and financial experience. Soc Sec Bull 1975;38(2):21–40.

14. Anderson O, Collette P, Feldman JJ. Changes in family medical care expenditures and voluntary health insurance. Cambridge: Harvard University Press, 1963.

15. U.S. Congressional Budget Office. Profile of health care coverage: The haves and have nots. Washington, D.C.: U.S. Government Printing Office, March 1979.

16. Kasper JA, Walden DC, Wilensky GR. Who are the uninsured? National Health Care Expenditure Study Data Preview 1. Hyattsville, Md.: National Center for Health Services Research, 1980.

17. Wersinger RP. Personal communication, December 23, 1980.

18. Winnard K. Personal communication, September 12, 1980.

19. Luft HS, Maerki SC. Duplicate health insurance coverage: Determinants of variation across states. Mimeo, 1980.

20. Price DN. Private Industry health insurance plans: Type of administration and insurer in 1974. Soc Sec Bull 1977;40(3):13–42.

21. Novotny A. Private industry health insurance plans: employment requirements for coverage in 1974. Soc Sec Bull 1977;40(3):28–33.

22. Kolodrubetz WW. Group health insurance coverage of full-time employees, 1972. Soc Sec Bull 1974;37(4):17–35.

23. Skolink AM. Twenty-five years of employee-benefit plans. Soc Sec Bull 1976;39(9):3–21.

24. U.S. Department of Labor and U.S. Department of Health, Education, and Welfare. 1978 employment and training report of the president. Washington, D.C.: U.S. Government Printing Office, 1978.

25. U.S. Bureau of the Census. Social indicators 1976. Washington, D.C.: U.S. Government Printing Office, 1977.

26. U.S. Bureau of the Census. Household and family characteristics, March 1977. Current population reports series P-20, no. 326, August 1978.

27. Deacon RW. Impact of Medicare on the use of medical services by disabled beneficiaries, 1972–1974. Health Care Financing Review 1979;1:39–54.

28. Health Care Financing Administration. Medicare: Inpatient use of short-stay hospitals, 1977. (DHEW publication no. (HCFA) 03022.)

29. National Center for Health Statistics. Utilization of short-stay hospitals: Annual summary for the United States, 1976. Vital and health statistics, series 13, no. 37. (DHEW publication no. (PHS) 78-1788, 1978.)

30. National Center for Health Services Research and National Center for Health Statistics. Health United States, 1978. (DHEW publication no. (PHS) 78-1232, 1978.)

31. Second opinion plan increases operations. (Associated Press). Peninsula Times Tribune (December 22, 1980), p. A-8.

32. Goldberg LG, Greenberg W. The HMO and its effects on competition. Washington, D.C.: Federal Trade Commission, Bureau of Economics, 1977.

33. Christianson J. Do HMOs stimulate beneficial competition? Excelsior, Minn.: InterStudy, 1978.

34. Luft HS. The selection and competitive effects of HMOs. Paper presented at American Public Health Association Annual Meeting, November 1979.

MEDICAL CARE
October 1981, Vol. XIX, No. 10

Mode of Payment as a Predictor of Health Status, Use of Health Services and Preventive Health Behavior:

A Report From the Los Angeles Health Survey*

ALFRED C. MARCUS, PH.D.†

Several issues relevant to Health Maintenance Organizations (HMOs) were examined in this article using data collected from the Los Angeles Health Survey. No support was found for the hypothesis that HMOs disproportionately attract people in poorer health. In fact, HMO members actually reported lower rates of acute or occasional illness and disability than fee-for-service subscribers (FFS). HMO members also reported lower total family incomes and educational status than FFS subscribers and they were less likely to report having a regular doctor. Few differences were found between the two groups in reports of using health services and practicing preventive health behavior. However, HMO members were more likely to report a recent physical examination—but only after taking into account the fact that they were less likely to have a regular doctor. It has been shown that these findings are consistent with recent evidence, and serve to amplify several key findings reported in other studies.

THE CONTINUING ESCALATION of health care expenditures is one of the most serious and controversial issues facing policymakers in the health care field. During the past decade expenditures for health care have increased at the rate of 12 per cent per annum, resulting in a *twofold increase* in health care costs within a 6-year period (1970–1976).[1,2] One of the advantages claimed by HMOs (particularly closed-panel HMOs) is that they will encourage reductions in the costs of medical care, especially costs associated with hospitalization. Although previous research is consistent with this hypothesis, different interpretations have been offered to explain the lower hospitalization rates among HMO members.[3–8] Advocates of HMOs argue that physicians in a prepaid group practice will be less likely to prescribe unnecessary medical treatments, since their fee is not dependent on the number of treatments they prescribe. In addition, since HMO physicians have a financial incentive for maintaining the health of their clients, it is believed they will encourage more preventive health care than FFS physicians.[3,5,6–9] Critics of HMO programs, on the other hand, claim that HMOs have a latent financial incentive to undertreat patients. Others have noted that the lack of a close doctor–patient relationship

---

*Data reported in this article were collected pursuant to grant number 5-R18-CA-18451, "Processes in Health Behavior and Cancer Control," awarded by the National Cancer Institute, DHEW, to the late Leo G. Reeder, Principal Investigator.

† School of Public Health, University of California, Los Angeles.

Send reprint requests to: Alfred C. Marcus, Director of Program Evaluation, Jonsson Comprehensive Cancer Center Division of Cancer Control, UCLA Center for Health Sciences, A3-307, Los Angeles, CA 90024.

0025-7079/81/1000/0995/$01.30 © J. B. Lippincott Co.

among members of a prepaid group practice may inhibit use of health services.[10] It has also been suggested that prepaid health insurance will attract people in greater need of health care, resulting in an "adverse self-selection bias" in HMOs.[11]

In this article HMO members have been compared with FFS subscribers on a number of measures related to these issues, including self-reports of illness and disability, preventive health behavior and use of physician services. These data were collected by personal interview as part of the Los Angeles Health Survey. Thus, they reflect the health care experiences of HMO and FFS members residing within a large, diverse metropolitan area. As will be shown below, our results corroborate and amplify findings from several recent studies.

Methods and Procedures

The Los Angeles Health Survey (LAHS) is a longitudinal survey of the general population of Los Angeles County that was initiated in 1974 by the late Leo G. Reeder. The overall research design incorporates both a trend and a panel survey. In the trend survey we have been comparing independent samples drawn from the same population at different points in time.[12] The main objective of the trend survey is to monitor aggregate population trends in health status, use of health services and adherence to medical advice. The panel survey permits analyses of these same variables using the individual (rather than the group) as the unit of analysis. In the panel survey respondents were reinterviewed every 6 to 8 weeks for approximately 1 year (1976–1977). The data reported below were obtained from the first panel interview conducted during the fall/winter of 1976. Analysis of the longitudinal date is currently in progress, and will be reported in a subsequent article.

The sampling design used in this survey is a three-stage random probability sample of Los Angeles County that was developed by the Institute for Social Science Research (ISSR) at UCLA.[13] The sampling frame contains approximately 20,000 computer-readable addresses sampled from the county on an area-probability basis. Samples drawn from the frame may be characterized as probabilities proportional to size (PPS) three-stage samples. The three stages use a computer sampling procedure which ensures that each housing unit in Los Angeles County has an equal probability of being selected.

Initially, 2,020 housing units were selected for the 1976 sample. However, 137 of these addresses were declared ineligible (e.g., vacancies, no dwelling unit, etc.), resulting in an eligible sample of 1,883. One adult respondent was then randomly selected from each household using the Kish[14] selection table. Of these 1,883 potential respondents, 1,210 (64.3 per cent) were successfully interviewed. Eighteen per cent of the eligible sample refused to participate, while another 10 per cent were coded as not at home (after a minimum of three callbacks). Nonresponse from all other sources totaled 8 per cent.

The LAHS interviews were conducted in the respondent's home, and were about an hour in length. The interview schedule covered a wide range of topics related to the health and health behavior of the respondents, including recent illnesses and disabilities, use of health services for these illnesses and preventive health behavior. Other questions asked about health insurance coverage. During the interview respondents were handed a card listing more than 20 health insurance companies and health plans; they were then asked whether they had health insurance. Those responding yes were asked to name their companies or health plans. This information was subsequently coded into five general categories of health insurance coverage: 1) no insurance; 2) Medi-Cal coverage; 3) Medicare coverage; 4) private fee-for-service insurance; and 5) prepaid HMO coverage. In the analyses reported below, comparisons have been made between the HMO and FFS groups. The HMO group originally contained 200 respondents. However, 10 of these people also belonged to a FFS health plan. To keep the two groups as distinct as possible, these 10 respondents were deleted from the analysis. Of the remaining 190 people in the HMO group, 81 per cent belonged to Kaiser, while another 14 per cent belonged to Ross–Loos. The FFS group contained 653 respondents, which excludes those belonging to an HMO, as well as those reporting Medicare or Medi-Cal coverage. The most popular insurance plans in the FFS category were Blue Cross (28 per cent) and Blue Shield (9 per cent).

The HMO–FFS comparisons reported in this article focus on the following dependent variables obtained from the household interviews:

Self-reported tendency to delay
Use of physician services for acute conditions (past 2 months)
Use of physician services for chronic conditions (past 2 months)
Physician recommended preventive health behavior
Recency of last reported physical examination
Cessation or reduction in smoking
Breast self-examination

The scoring procedures for these variables, as well as their mean scores, are presented in Table 1. The actual questionnaire items used to create these variables are presented in Appendix A.

The nine dependent variables shown in Table 1 were examined using multiple regression. The "mode of payment" variable was coded in dummy variable format

TABLE 1. Variable Names, Scoring Procedures and Mean Scores for the Dependent Variables

Variables	Scoring	$\bar{X}$	N*
Self-reported tendency to delay	1 = respondent says he/she is a "long delayer" 0 = respondent says he/she delays a moderate or short amount of time	0.22	843
Use of physician services for acute conditions	Proportion of acute problems reported by respondent that resulted in doctor contact (past 2 months)	0.51	248
Use of physician services for chronic conditions	Proportion of chronic problems reported by respondent that resulted in doctor contact (past 2 months)	0.32	410
Physician recommended preventive health behavior (PHB)	1 = respondent says physician recommends some form of PHB 0 = respondent reports no PHB is being recommended	0.23	843
General physical examinations (GPE)	Number of years since person last reported a GPE. Scores ranged from 1–8+ years.	2.63	843
Cessation of smoking (among those who smoked for at least 1 year)	1 = respondent says he/she quit smoking cigarettes 0 = respondent reports he/she still smokes	0.34	491
Reduction in smoking (among current smokers)	1 = respondent reports reduction in smoking when compared to "recent years" 0 = no reduction reported	0.29	323
Breast self-examination (BSE) (among women)	1 = respondent says she practices BSE 0 = no BSE is reported	0.61	442
Frequency of BSE	1 = respondent reports BSE monthly or more often 0 = no report of monthly BSE	0.39	442

* Group size will vary depending on the item. The base number of 843 represents the combined categories of HMO (N = 190) and FFS (N = 653).

(HMO = 1; FFS = 0), and entered as one of the predictor variables. Other predictor variables include a host of demographic and health status measures that serve as control variables in this analysis (e.g., age, sex, educational and occupational ranking, total family income, recent illness history, etc.). In the interest of space, we have only reported the regression coefficients obtained for the "mode of payment" variable. However, the regression coefficients for the other predictor variables are available for review in Appendix B.

As shown in Table 1, virtually all dependent variables are scored as either dichotomies or proportions (i.e., the scores ranged from 0 to 1). For such variables, the size of the (unstandardized) regression coefficient for "mode of payment" represents the difference in proportions between the HMO and FFS groups—estimated after all other variables in the equation have been controlled statistically. Given the coding of the mode of payment variable, a positive regression coefficient would indicate that the HMO group scored higher on that dependent variable than the FFS group, while a negative coefficient would indicate the opposite. Thus, if the FFS group scored at the mean for "general tendency to delay" ($\overline{X} = 0.22$), and the regression coefficient for mode of payment was −0.02, then the proportion for the HMO group would be 0.02 lower than the FFS group (or 0.20). Of course, if the regression coefficient had been positive rather than negative, then the HMO group would have scored 0.02 higher (or 0.24).

Results

Before comparing the HMO and FFS groups in their (self-reported) use of health services and preventive health behavior, we should first determine their comparability on other related factors, including differences in sociodemographic characteristics and recent morbidity. As shown in Table 2, HMO members had lower education and total family incomes, and they were more likely to be black and Hispanic. HMO members were also more likely to have children (under the age of 18) living in the household. There were no differences between the two groups in age, sex or occupational ranking.

Given this demographic profile, one might expect HMO members to be at slightly greater risk for health care than FFS subscribers. However, HMO members actually reported less acute illness and disability within 2 months of the interview (see Table 3). There were no differences in reports of chronic illness or in self-rated health status.

The lower rates of acute illness reported by HMO members occurred primarily in symptoms of the respiratory tract. As shown in Table 4, 7.9 per cent of the HMO group reported acute respiratory symptoms during the past 2 months, compared with 16.8 per cent of the FFS group. Differences were most evident in reports of cold and flu symptoms. Eight HMO members reported cold symptoms (4.2 per cent), compared with 56 FFS subscribers (8.6 per cent). Similarly, only five HMO members reported the flu (2.6 per cent), compared with 48 FFS subscribers (7.4 per cent). Aside from these two differences, however, the HMO and FFS groups reported very similar rates of acute illness. The distribution of chronic symptoms also appears comparable, with the largest differences occurring for symptoms of the nervous system and mental health.

Use of Physician Services

Most studies have shown a tendency for HMO members to report lower rates of hospitalization and surgeries,[3-8,15-25] but higher rates of ambulatory and physician visits.[3,4,6-8,26-33] Since our measures of use deal exclusively with physician visits, we expected to find higher reporting levels in the HMO group. Contrary to expectation, however, this hypothesis was not sup-

TABLE 2. Cross-tabulation of Demographic Characteristics by Mode of Payment

	HMO (N = 190)	FFS (N = 653)	p ≤
Sex			
Female	52.7	52.4	NS
Male	47.3	47.6	
Age			
18–29	30.7	33.6	
30–39	24.4	21.8	
40–49	20.1	18.9	NS
50–59	14.9	19.4	
60+	10.0	6.3	
Education			
Less High School	28.3	18.8	
High School	24.6	26.3	
Some college	22.9	31.1	.01
College grad and above	24.1	23.8	
Occupation			
Operatives	12.4	13.0	
Crafts	6.0	10.1	
Sales/clerical	27.2	31.7	NS
Managerial	9.4	10.4	
Professional/technical	20.2	20.9	
Income			
< 5,000	7.4	2.9	
5,000–9,000	16.6	16.2	
10,000–13,999	22.1	19.7	
14,000–19,999	24.4	24.9	.05
20,000–24,999	13.2	14.3	
25,000+	16.3	22.0	
Race			
White/Anglo	59.8	75.2	
Hispanic	22.4	17.6	.001
Black	17.8	7.2	
Children in household			
Yes	56.2	47.8	
No	43.8	52.2	.05

NS: not significant.

ported. To illustrate this point, we have presented the regression coefficients for "mode of payment" in nine separate analyses. As shown in Table 5(A), the HMO and FFS groups did not differ statistically in their self-assessed tendency to delay, or in the proportion of acute and chronic health problems that resulted in a doctor contact within 2 months of the interview.

In Table 5(B) we have presented comparable results obtained for several particular symptoms. These measures were obtained for all symptom codes that satisfied two nominal criteria: the number of cases was ≥ 50, and the distribution of the proportion had no more than 75 per cent of the cases in the modal category. Six symptom codes satisfied these criteria: acute respiratory problems, acute flu symptoms, acute musculoskeletal problems, chronic cardiovascular problems—particularly high blood pressure—and chronic musculoskeletal problems. As shown, HMO members reported fewer doctor contacts for acute respiratory and chronic musculoskeletal symptoms, but more frequent doctor contacts for acute

TABLE 3. Mean Scores for HMO Members and FFS Subscribers in Self-Reports of Illness, Disability and Health Status

	$\overline{X}$ Scores	
	HMO (N = 190)	FFS (N = 653)
Average number of acute illnesses reported past 2 months	0.23	0.39*
Average number of chronic illnesses reported	0.68	0.70
Proportion reporting bed days past 2 months	0.11	0.23*
Proportion reporting reduced activity days past 2 months (not counting bed days)	0.12	0.28*
Self-rated health status (proportion reporting excellent or good health)	0.83	0.86

* $p < .001$ (F-test).

musculoskeletal symptoms and symptoms of the cardiovascular system. However, since these differences are based on a smaller number of cases, the regression coefficients would have to be much larger to attain statistical significance (i.e., none of the regression coefficients reported in Table 5 is statistically different from 0). Thus, the picture that begins to emerge from our data is one of similarity between the two groups in their use of physician services.

This conclusion is reinforced even further when we compare the two groups in terms of their "use/disability ratios." The use/disability ratio is an aggregate index for groups that measures the number of doctor contacts per 100 disability days (bed days + reduced activity days).[12,34,35] Thirty-four people from the HMO group reported 263 disability days and 26 "initial" doctor contacts. Accordingly, there were approximately 9.9 doctor contacts per 100 disability days in the HMO group. In the FFS group 246 people reported 2,084 disability days and 183 doctor contacts, for a use/disability ratio of nearly 8.8 doctor contacts per 100 disability days. In other words, the two groups differed by slightly more than one doctor contact per 100 disability days—a very small difference indeed.‡

Preventive Health Behavior

One frequently hears that prepaid health care will create a financial incentive for the practice of preventive medicine. The available evidence, however, is somewhat inconsistent on this point. Some studies have shown that HMO members receive more preventive health care than FFS subscribers; other studies have found little or no differences between modes of payment.[9] Our results are much more consistent with the latter group of studies. As shown in Table 6, there were no differences in reports of receiving preventive health advice from physicians, or in the recency of physical examinations. Similarly, there were no differences in the percentages of exsmokers or in the percentage of current smokers who report smoking less. Among the women, there were no differences in reports of BSE.

Identification With a Regular Doctor

During the interview all respondents were asked whether "there is one person you usually see" for your health care and whether "this person is a medical doctor."

---

‡ We might note that the FFS group had a somewhat larger proportion of people reporting long-term disability—particularly people reporting 30 or more restricted activity days during the past 2 months. For example, only 3.2 per cent of the HMO group reported 30+ days of restricted activity, compared with 7.9 per cent in the FFS group. However, the proportion reporting 25+ restricted activity days was about the same for the two groups. To adjust for the effects of these "outliers," we truncated the individual reports of disability at 25+ days and recalculated the use/disability ratios. Interestingly, these adjusted rates were virtually identical for the two groups: 11.45 doctor contacts/100 disability days for the HMO group, compared with 11.74 doctor contacts for the FFS group.

TABLE 4. Frequency and Per Cent Reporting Particular Symptom Codes by Mode of Payment

	No. and Per Cent Reported as Acute Symptoms				No. and Per Cent Reported as Chronic Symptoms			
	HMO (N = 190)		FFS (N = 653)		HMO (N = 190)		FFS (N = 653)	
Symptom Categories*	N	%	N	%	N	%	N	%
Nonspecific (0.0–0.49)	0	0.00	6	0.92	4	2.11	10	1.53
Nervous system (50.0–99.0)	0	0.00	4	0.61	12	6.32	21	3.21
Skin, nails and hair (100–199)	1	0.53	15	2.30	2	1.05	21	3.21
Cardiovascular and lymphatic systems (200–299)	0	0.00	5	0.76	19	10.00	62	9.49
Respiratory system (300–399)	16	7.89	110	16.85	23	12.10	77	11.79
Musculoskeletal system (400–499)	17	8.95	60	9.19	28	14.74	93	14.24
Digestive system (500–599)	6	3.16	23	3.52	11	5.79	42	6.43
Urinary system (600–629)	1	0.53	6	0.92	2	1.05	11	1.68
Reproductive system (630–699)	1	0.53	8	1.23	5	2.63	20	3.06
Eyes and ears (700–799)	3	1.58	4	0.61	7	3.68	19	2.91
Mental health (800–899)	0	0.00	0	0.00	5	2.63	4	0.61

* Symptoms were coded using the National Ambulatory Medical Care Survey Symptom Classification. The numbers in parentheses denote the code values.

Fifty percent of the HMO members reported having a regular doctor, compared with 75 per cent for the FFS group ($p < 0.001$). Recently, Scitovsky and associates[10] have shown that the lack of a close doctor–patient relationship may hinder use of physician services. Thus, one might hypothesize that doctor visits for HMO members will increase, relative to FFS, when differences in having a regular doctor are taken into account. Given the coding for the "mode of payment" variable, this hypothesis would predict that the regression coefficient for mode of payment will increase when "regular doctor" is added to the regression. Interestingly, we found one such case among our dependent variables. As reported earlier in Table 6, when we examined the number of years since the person last reported a physical examination, the regression coefficient for "mode of payment" was -0.22 ($F = 1.32$, not significant). When "regular doctor" was added to the equation, the regression coefficient for mode of payment increased to -0.47 ($F = 5.55$, $p < 0.05$). Thus, we found that HMO members reported a more recent physical examination—but only after taking into account the fact that they were less likely to have a regular doctor.

As noted in Appendix A, use of physician services for a physical examination was analyzed in several different ways. The finding reported in the previous paragraph pertains to the measure based on number of years (trimmed at 8+ years to reduce skewness). When we examined the logarithm of the original variable, the same exact pattern emerged. That is, before adding "regular doctor" to the regression, the regression coefficient for mode of payment was only 0.04 ($F = 0.40$, not significant). However, when regular doctor was added as a control variable, the regression coeffi-

TABLE 5. Regression Coefficients (β) and Sample Size (N) for "Mode of Payment" and Self-Reported Use of Health Services*

	β	N
A. Use of health services—general measures		
Self-reported tendency to delay (proportion reporting being long delayers)	−0.02	(840)
Proportion of acute problems (past 2 months) resulting in doctor contact	0.00	(247)
Proportion of chronic problems (past 2 months) resulting in doctor contact	−0.05	(409)
B. Use of health services—particular symptoms†		
Proportion of acute respiratory problems (300–399) resulting in doctor contact (past 2 months)	−0.07	(125)
Proportion of acute flu symptoms (313) resulting in doctor contact (past 2 months)	−0.14	(53)
Proportion of acute musculoskeletal problems (400–499) resulting in doctor contact (past 2 months)	0.05	(71)
Proportion of cardiovascular problems (200–299) resulting in doctor contact (past 2 months)	0.01	(81)
Proportion of high blood pressure problems (205) resulting in doctor contact (past 2 months)	0.14	(51)
Proportion of chronic musculoskeletal problems (400–499) resulting in doctor contact (past 2 months)	−0.13	(121)

* The "mode of payment" variable was coded 1 = HMO, 0 = FFS. The (unstandardized) regression coefficients represent differences in mean scores between the two groups after taking into account all other variables in the equation, including: respondent's age, sex, total family income, marital status (1 = married, 0 = not married), race/ethnicity (1 = White/Anglo, 0 = Other), educational and occupational ranking, reduced activity and bed days (past 2 months), number of acute illnesses (past 2 months), number of chronic illnesses and self-rated health status. The number in parentheses (N) is the size of the group being analyzed. Group size will vary depending on the item.

† All symptoms were coded using the National Ambulatory Medical Care Survey System Classification. The numbers in parentheses denote the code values.

cient for mode of payment increased to 0.12 ($F = 3.70$, $p < 0.06$). Additional measures included rates of people reporting a physical examination within 1, 2, 3, 4, and 5 years of the interview. For rates that extended 4 years and beyond, the pattern described above was replicated. Thus, when predicting whether the person had a physical examination within 4 years of the interview, the regression coefficient for mode of payment was only 0.04 ($F = 1.46$, not significant). However, after adding "regular doctor" to the regression, the regression coefficient for mode of payment increased to 0.07 ($F = 4.91$, $p < 0.05$). Interestingly, this same effect was *not* in evidence when examining 1-, 2-, and 3-year rates of physical examinations. This pattern suggests that prepaid group practice, when compared with comparable FFS practice, may have a lower proportion of people who delay getting a physical examination for a relatively long period of time (e.g., four or more years). However, there appears to be little difference between the two groups in the proportion of people receiving a more recent physical examination (e.g., within the past 2 or 3 years).

Summary and Discussion of Results

Perhaps our findings can best be organized and discussed in relation to several key policy issues relevant to HMO legislation. For example, one such issue is whether HMOs disproportionately enroll people at "greater risk." In a recent study of this hypothesis, Berki and associates[11,36] distinguished between two forms of

TABLE 6. Regression Coefficients (β) and Sample Size (N) for "Mode of Payment" and Self-Reported Preventive Health Behavior*

Preventive Health Behavior	β	N
Proportion reporting preventive health advice from doctor.	0.01	(839)
No. of years since last reported physical examination. (Scores trimmed at 8+ years).	−0.22	(839)
Proportion stopped smoking (among those who smoked for at least a year).	−0.03	(490)
Proportion smoking less (among current smokers).	−0.03	(322)
Proportion practicing breast self-examination (BSE) (among women)	−0.04	(440)
Proportion practicing monthly BSE (among women).	−0.01	(440)

* The "mode of payment" variable was coded 1 = HMO, 0 = FFS. The (unstandardized) regression coefficients represent differences in mean scores between the two groups after taking into account all other variables in the equation, including respondent's age, sex, total family income, marital status (1 = married, 0 = not married), race/ethnicity (1 = White/Anglo, 0 = Other), educational and occupational ranking, reduced activity and bed days (past 2 months), number of acute illnesses (past 2 months), number of chronic illnesses and self-rated health status. The number in parentheses (N) is the size of the group being analyzed. Group size varies depending on the item.

risk—one related to health, the other to costs (for an earlier treatment of this hypothesis see Bashur and Metzner[37]). According to the financial-risk hypothesis, HMOs will enroll more people vulnerable to financial loss than FFS, resulting in an "adverse economic risk self-selection factor" in HMOs. Data from previous studies are generally consistent with the financial vulnerability hypothesis.[11,26,38,39] Our data also support this proposition, since HMO members had lower total family incomes than FFS subscribers. Additionally, we found that HMO members were more likely to have children living in the household. This finding supports previous research showing that "HMOs attract families expected to exhibit greater demand for care that is associated with a younger and expanding life stage"[11] (p. 607). However, HMO enrollees did not report different rates of chronic illness or poorer overall health. In fact, the only statistical difference we found had HMO members reporting less acute illness and disability than FFS subscribers—particularly cold and flu symptoms. Consequently, we have no evidence to support the health–risk hypothesis that HMOs will enroll more people in poor health. Although a few early studies reported data in support of the health-risk hypothesis,[6,37,40,41] much of the recent evidence is consistent with our findings of no adverse health risk among HMO members.[11,15,26,36,38,39,42,43]

Another key issue is whether the HMO and FFS groups use health sevices in a comparable fashion. Most studies have shown a tendency for HMO members to report higher rates of ambulatory and physician visits than FFS.[3,4,6,18,26–32] However, we found no differences in reports of using physician services for either acute or chronic illness. As noted previously, our indicators of use measure the proportion of health problems that resulted in a doctor contact (within the previous 2 months). Consequently, these measures primarily reflect *entry* into the health care system via a physician visit. They do not reflect use of services for follow-up visits. Thus, if HMO members are more likely to have follow-up visits to the doctor, then this difference would not be detected in this analysis.

One clear difference did emerge between the two groups concerning their identification with a regular doctor. HMO enrollees were much less likely to report a regular doctor than FFS subscribers. Several studies have shown a positive association between use of services and having a regular physician.[10,44] For example, Scitovsky and associates[10] have recently compared use of services by Kaiser members with that of a group practice clinic plan that was predominantly fee-for-service.§ They found that the Kaiser sample had lower rates of use than the clinic sample, which they attributed to the smaller percentage of Kaiser members reporting a regular doctor.

The data reported by Scitovsky et al. were obtained from a sample of faculty and staff at Stanford University. Consequently, the generality of their results might be questioned. It is of interest to note, therefore, that we uncovered one finding that is consistent with their results. That is, we found that reports of using physician services for physical examinations increased for HMO members, relative to FFS, when initial differences in having a regular doctor were taken into account. Thus, one might speculate, as Scitovsky and associates[10] have done, that "use of physician services may have been held down by the relative lack of a close patient–physician relationship" (p. 15) among Kaiser members. On the other hand, we should not forget that the overall size of this effect was quite modest in our data. Future research would be well-spent examining this issue in more detail.

Supporters of HMO legislation have maintained that prepaid health insurance will encourage the practice of preventive health care. Although previous research is generally consistent with this proposition,[18,42,45–47] several studies have reported mixed or contradictory results.[19,24,48] Luft[9] has recently reviewed this literature and concluded that "the greater number of preventive services among HMO enrollees seem merely to result from the lower price patients face to ambulatory visits [in HMOs]" (p. 141). In other words, Luft has argued that HMO members receive more preventive health care not because HMO physicians are more committed to the practice of preventive medicine, but because preventive health services are generally less expensive to the consumer in an HMO. Luft's conclusion is particularly noteworthy, since it discounts the popular notion that prepaid health care creates a financial incentive for physicians to practice more preventive medicine. Interestingly, our data tend to corroborate Luft's reading of the literature. For example, if HMO physicians were operating strictly from the perspective of their financial interests (i.e., if they were primarily interested in reducing future treatment costs for the HMO) then we would expect them to offer more preventive health advice than FFS physicians. However, the HMO and FFS groups did not differ in their reports of receiving preventive health advice from their doctors (e.g., stop smoking, exercise more often). Similarly, no differences were found in two health behaviors that might be expected to reduce future treatment costs for the HMO: cancer screening behavior (BSE) and cessation or reduction of smoking. On the other hand, differences were found in a behavior consistent with Luft's hypothesis. Virtually all of the HMO members had prepaid insurance covering "health maintenance examinations," while a much smaller proportion of the FFS group would be expected to have comparable coverage. As Luft would predict, HMO members reported more recent physical examinations than FFS subscribers. However, as noted previously, we observed this difference only after taking into account the fact that HMO members were less likely to have a regular doctor. This latter finding suggests that HMO members might be more likely to take advantage of their preventive health coverage when they have a close doctor–patient relationship.

---

§ More specifically, the clinic plan examined by Scitovsky and associates differed from the Kaiser plan in that 1) care was provided by a predominantly fee-for-service group practice whose prepaid patients accounted for only about 15 per cent of its annual gross revenue; 2) the group was not at risk for the hospital costs of their prepaid patients; and 3) the physicians often did not know whether the patient was prepaid or fee-for-service.

As Andersen and Aday[44] note, the importance of having a regular source of care underscores the potential value of HMO legislation. Yet, the advantages associated with closed-panel HMOs (e.g., providing a regular place for health care) may be confounded by certain obstacles that limit or hinder use of services. One such obstacle is the tendency for people without a regular doctor to be enrolled in closed-panel HMOs. What remains unclear about this finding is the extent to which closed-panel HMOs discourage identification with a regular physician, as compared to people simply enrolling in closed-panel HMOs because they have no regular doctor. Several studies have shown that enrollment in closed-panel group practice is more likely to occur among people having no regular doctor.[8,11,26,38,41,43] However, FFS physicians may also have a financial incentive to establish a long-term relationship with their clients that is missing in a closed-panel HMO.[10] Thus, future research might examine whether the absence of a regular doctor is basically an "input" characteristic of the enrollment population of HMOs, or is primarily a "process" variable associated with the formal organization of closed-panel group practice. Such information should be of considerable benefit to health administrators interested in improving the operation and management of existing HMOs.

References

1. Roy WR. Background and plan of conference. In: Effects of payment mechanism on the health care delivery system. W. Roy, ed., Bethesda, Md.: U.S. Department of Health Education and Welfare, National Center for Health Services Research, 1978. (DHEW publication no. 78-3227.)

2. Public Health Service. Health United States 1976–1977. U.S. Department of Health, Education, and Welfare, Public Health Service, Health Resources Administration, 1977. (DHEW publication no. (HRA) 77-1233.)

3. Chassin MR. The containment of hospital costs: a strategic assessment. Med Care 1978; 16(Oct Suppl):14.

4. Luft HS. How do health maintenance organizations achieve their "savings"? N Engl J Med 1978;298:1336.

5. Klarman HE. Analysis of the HMO proposal—its assumptions, implications and prospects. In: Health organizations: a reconfiguration of the health services system. Chicago: University of Chicago, Center for Health Administration Studies, 1971:24–38.

6. Roemer ML, Shonick W. HMO performance: the recent evidence. Milbank Mem Fund Q 1973;51:271.

7. Wolinsky FD. The performance of health maintenance organizations: an analytic review. Milbank Mem Fund Q 1980;58:537.

8. Luft HS. Assessing the evidence of HMO performance. Milbank Mem Fund Q 1980;58:501.

9. Luft HS. Why do HMOs seem to provide more health maintenance services? Milbank Mem Fund Q 1978;56:140.

10. Scitovsky AA, Benham L, McCall N. Use of physician services under two prepaid plans. Med Care 1979;17:441.

11. Berki SE, Ashcraft MLF. HMO enrollment: who joins what and why: a review of the literature. Milbank Mem Fund Q 1980;58:588.

12. Marcus AC, Reeder LG, Jordan LA, Seeman TE. Monitoring health status, access to health care and compliance behavior in a large urban community: a report from the Los Angeles Health Survey. Med Care 1980;18:253.

13. Sumner J. The 1976 LAMAS frame and master sample: technical description. Los Angeles: Institute for Social Science Research, University of California, Los Angeles, 1976.

14. Kish L. Survey sampling. New York: Wiley, 1965.

15. Mott FP, Hastings JE, Barclay AB. Prepaid group practice in Sault Ste. Marie, Ontario: part III: evidence from the household survey. Med Care 1973;11:173.

16. Wersinger R, Roghmann KJ, Gavett JW, Wells SM. Inpatient hospital utilization in three prepaid comprehensive health care plans compared with a regular Blue Cross plan. Med Care 1976;14:721.

17. Perkoff GT, Kahn L, Haas PJ. The effects of an experimental prepaid group practice on medical care utilization and cost. Med Care 1976;14:432.

18. Hastings JE, Mott FD, Barclay A, Hewitt D. Prepaid group practice in Sault Ste. Marie, Ontario: part I: analysis of utilization records. Med Care 1973;11:91.

19. Gaus CR, Cooper BS, Hirschman CG. Contrasts in HMO and fee-for-service performance. Soc Sec Bull 1976;39:3.

20. Riedel DC, Walden DC, Singsen AG, et al. Federal employees health benefits program: utilization study. Washington, D.C.: National Center for Health Services Research, 1975.

21. Payne BC, Lyons TF. Method of evaluating and improving personal medical care quality. Ann Arbor: University of Michigan, 1972.

22. Sullivan R. Health maintenance groups take growing role as resistance eases. New York Times, May 21, 1978.

23. McCarthy E, Widmer G. Effects of screening by consultants on recommended elective surgical procedures. N Engl J Med 1974;291:1331.

24. DeFriese GH. On paying the fiddler to change the tune: further evidence from Ontario regarding the impact of universal health insurance on the organization and patterns of medical practice. Milbank Mem Fund Q 1975;53:117.

25. Choi J, Ries P. Sociodemographic and health characteristics of persons by private health insurance coverage and type of plan: United States, 1975. Advancedata, Vital and Health Statistics, National Center for Health Statistics, 32: August 22, 1978.

26. Ashcraft M, Penchansky R, Berki SE, Fortus RS. Expectations and experiences of HMO enrollees after one year: an analysis of satisfaction, utilization and costs. Med Care 1978;16:14.

27. Broida JH, Lerner M, Lohrenz FN, Wenzel FJ. Impact of membership in an enrolled, prepaid population on utilization of health services in a group practice. N Engl J Med 1975;292:780.

28. Garfield SR, Collen MI, Feldman R, et al. Evaluation of an ambulatory medical care delivery system. N Engl J Med 1976;294:426.

29. Levine DM, Morlock LL, Mushlin AL, et al. The role of new health practitioners in a prepaid group practice: provider differences in process and outcomes of medical care. Med Care 1976;14:326.

30. Record JC, Hurtado AV, Obannon JE. Quality of physician assistant performance at a health maintenance organization. In: Bliss AA, Cohen EP, eds. The new health professionals. Germantown, Md.: Aspen Publications, 1977.

31. Sackett DL, Spitzer WO, Gent M, et al. The Burlington randomization trial of the nurse practitioner: Health outcomes of patients. Ann Intern Med 1974;80:137.

32. Saward EW, Greenlick MR. Health policy and the HMO. Milbank Mem Fund Q 1972;50:147.

33. Blumberg MS. Health status and health care use by type of private health coverage. Milbank Mem Fund Q 1980;58:633.

34. Aday LA. Economic and noneconomic barriers to the use of needed medical services. Med Care 1975;13:447.

35. Andersen R. Health status indices and access to medical care. Am J Public Health 1978;68:458.

36. Berki SE, Ashcraft ML, Penchansky R, Fortus RS. Enrollment choice in a multi-HMO setting: the roles of health risk, financial vulnerability and access to care. Med Care 1977;15:95.

37. Bashshur RL, Metzner CA. Vulnerability to risk and awareness of dual choice of health insurance plan. Health Serv Res 1970;5:106.

38. Berki SE, Penchansky R, Fortus RS, et al. Enrollment choices in different types of HMOs: A multivariate approach. Med Care 1978;16:682.

39. Scitovsky A, McCall N, Benham L. Factors affecting the choices between two prepaid plans. Med Care 1978;16:660.

40. Gaus C. Who enrolls in a prepaid group practice: the Columbia experience. Johns Hopkins Med 1971;128:9.

41. Tessler R, Mechanic D. Factors affecting the choice between prepaid group practice and alternative insurance programs. Milbank Mem Fund Q 1975;53:149.

42. Berki SE, Ashcraft M, Penchansky R, Fortus R. Health concern, HMO enrollment and preventive care use. J Community Health 1977;3:3.

43. Roghmann KJ, Gavett JW, Sorensen AA, et al. Choosing between three health care plans: the Rochester experience. Rochester, N.Y.: Department of Pediatrics, University of Rochester, 1974. (Working paper no. 6.)

44. Andersen R, Aday LA. Access to medical care in the U.S.: realized a potential. Med Care 1978;16:533.

45. Donabedian A. An evaluation of prepaid group practice. Inquiry 1969;6:3.

46. Hetherington RW, Hopkins CE, Roemer MI. Health insurance plans: promise and performance. New York: Wiley-Interscience, 1975.

47. Slesinger DP, Tessler RC, Mechanic D. The effects of social characteristics on the utilization of preventive medical services in contrasting health care programs. Med Care 1976;14:392.

48. Diehr P, Richardson WC, Drucker WF, et al. Utilization: ambulatory and hospital. Chapter 2 of the Seattle Prepaid Health Care Project: comparison of health services delivery. Seattle: University of Washington, 1976.

Appendix A

All variables examined in this analysis were obtained from the 1976 household interview conducted as part of the Los Angeles Health Survey. The individual questionnaire items used as dependent variables are presented below.

Physical Examinations: "When was the last time you had a general physical examination or check-up (not because of any health problem or injury)?"

Answers to this question were originally coded in either days, weeks, months or years, depending on what was most convenient for the respondent. These responses were then converted into an index measuring the number of years since the respondent reported having a general physical examination. The original codes for this index ranged from 1 to more than 30 years, and were highly skewed. To reduce the skew in this variable, scores were "trimmed" at 8+ years (skew = 1.37). Additionally, the logarithm of this variable was examined (skew = 0.81), as were several dichotomous variables that represent different rates of physical examinations. Five such dichotomies were examined, representing the proportion of people reporting a physical examination within 1, 2, 3, 4 and 5 years of the interview. For the most part, the same results were obtained regardless of which measure was used.

Acute Illness: "Think back to the past 2 months; were you sick or did you have any accidents or injuries at any time during the past two months? Please do not include any recurrent or continuing health problem that you may have all of the time. We are interested only in acute or occasional illness or injuries."

Respondents answering "yes" to this question were asked to describe their acute illnesses (symptomatology and/or diagnosis), and whether they contacted a doctor about each problem. Up to three separate illnesses/injuries were recorded by the interviewers.

Chronic Illness: "Do you have a chronic—that is, a recurring or continuing—health problem, a health problem that you have all the time or makes you ill from time to time?"

The same information was obtained from this question as the one concerning acute illness. That is, respondents were asked to describe their chronic problems (up to three mentions), and the length of time since they last contacted a doctor about each problem.

Proportion of Acute/Chronic Problems Resulting in Doctor Contact

The acute and chronic questions listed above asked whether the respondent had "seen or made contact" with a doctor about each problem that was mentioned, and how long ago this contact occurred. This information was used to create two indexes measuring use of physician services: the proportion of acute problems that resulted in a doctor contact within 2 months of the interview and the proportion of chronic problems that resulted in a doctor contact within the same 2-month period.*

---

* Use of physician services for chronic illness was also examined using a 6-month and 1-year recall period. All three versions produced comparable results. We have focused on the version using a 2-month recall period because it was also used in several other questions that play a key role in this analysis (e.g., number of bed days and reduced activity days).

Another question asked about the respondent's general tendency to delay using health services:

> "When you think there is something wrong with your health, would you say you:
> Wait as long as possible before seeing a doctor;
> Put off going to the doctor for a short time; or
> Try to see a doctor immediately?"

Two questions asked about restricted activity during the past 2 months:

> *Bed Days:* "During the past 2 months, did you stay in bed all or part of any day because of any illness or injury?"
>
> *Reduced Activity Days:* "(Besides the days in bed) were there any other days during the past 2 months that you had to cut down on things you usually do because of poor health?"

Respondents were also asked to rate their health on a 4-point scale:

> How would you rate your health overall; would you say your health is:
> excellent;
> good;
> fair; or
> poor?"

Another question asked whether their doctors recommended that they change their behavior or life-style:

> "We have talked about some illnesses and health problems. Now I'd like to find out about any health problems, treatments or medicines that you may be taking or doing. Has a doctor prescribed or recommended that you take medicines, medical treatments or change your way of living in such areas as smoking, special diet, exercise, drinking or some other area?"

Respondents answering "yes" to this question were asked to describe these recommendations. Up to three mentions were recorded by the interviewers. The responses were then coded into 15 categories grouped according to whether they referred to behavioral/life-style changes or medications/treatments and other therapeutic devices (e.g., life-support devices, orthopedic braces, etc.). The specific categories depicting behavioral/life-style changes included: eating or avoiding specified foods or following a special diet, reducing activity, increasing activity or exercise, reducing consumption of cigarettes and alcoholic beverages, changes in work habits and getting more rest/relaxation. The common denominator among these recommendations is that they all exemplify some form of "preventive health advice."

Finally, additional questions asked about smoking behavior and self-examination of the breast. All respondents who reported smoking cigarettes for at least one year were asked whether they still smoke cigarettes, and (if yes), whether they are smoking more, less or about the same amount "as in recent years." The women were also asked whether they examine their brests for lumps, and (if yes) how frequently they perform breast self-examination.

Appendix B

Appendix B contains the scoring procedures and standardized regression coefficients for all predictor variables reported in Tables 5 and 6.

TABLE B1. Scoring Procedures for Selected Predictor Variables

Predictor Variables	Scoring Procedures
"Mode of payment"	1 = HMO; 0 = FFS
Sex	1 = female; 0 = male
Marital status	1 = married; 0 = not married
Socioeconomic index (SEI)	Duncan's Socioeconomic Index, where low score = low SEI, and high score = high SEI. Scores ranged from 001 to 996.
Race/ethnicity (self-reported)	1 = white/Anglo; 0 = other
Income	15-point scale, ranging from <$5,000 to $40,000+
Bed days	1 = 1+ bed days; 0 = no bed days
Reduced activity days	1 = 1+ reduced activity days; 0 = no reduced activity days
Self-rated health status	1 = excellent, 2 = good, 3 = fair, 4 = poor
Number of chronic health problems	1–3+
Number of acute health problems	1–3+

TABLE B2. Standardized Regression Coefficients and Sample Size (N) for Variables used in Predicting Self-Reported Use of Physician Services (General Measures)*

	Dependent Variables—Use of Physician Services		
Predictor Variables	Self-Reported Tendency to Delay (N = 840)	Proportion of Acute Problems Resulting in Doctor Contact (N = 247)	Proportion of Chronic Problems Resulting in Doctor Contact (N = 409)
"Mode of payment"	−0.02	0.00	−0.05
Age	−0.02	0.07	0.08
Sex	−0.05	0.02	0.13†
Marital status	0.04	−0.05	0.00
Socioeconomic index	0.12†	−0.09	0.00
Race/ethnicity	0.06	−0.06	−0.06
Income	−0.04	−0.04	0.07
Bed days	−0.07	−0.01	0.04
Reduced activity days	0.05	0.03	0.11‡
Self-rated health	0.10‡	0.07	0.20§
Chronic problems	0.01	−0.05	−0.13†
Acute problems	0.04	0.09	0.08
R	0.19	0.22	0.30

* The standardized regression coefficients reported in Table B2 were obtained from the same equations which produced the unstandardized regression coefficients reported in Table 5.

† $p < 0.01$.

‡ $p < 0.05$.

§ $p < 0.001$.

TABLE B3. Standardized Regression Coefficients and Sample Size (N) for Variables Used in Predicting Self-Reported Use of Physician Services (Specific Measures)*

Predictor Variables	Dependent Variables—Use of Physician Services					
	Acute Respiratory (300–399) (N = 125)	Acute Flu (313) (N = 53)	Acute Musculoskeletal (400–499) (N = 71)	Cardiovascular (200–299) (N = 81)	High Blood Pressure (205) (N = 51)	Chronic Musculoskeletal (400–499) (N = 121)
"Mode of payment"	−0.05	−0.09	0.04	0.02	0.12	−0.13
Age	0.00	−0.18	−0.03	0.09	−0.04	−0.06
Sex	−0.02	−0.11	−0.05	0.11	0.10	0.02
Marital status	0.09	0.12	0.00	−0.05	−0.15	0.03
Socioeconomic status	0.12	0.18	0.06	−0.18	−0.06	−0.08
Race/ethnicity	0.09	0.13	0.04	−0.16	−0.31	0.00
Income	−0.07	−0.03	−0.03	0.13	0.23	0.02
Bed days	−0.06	−0.04	−0.07	−0.21	−0.05	0.08
Reduced activity days	0.04	0.15	−0.06	0.16	0.14	0.14
Self-rated health status	−0.16	−0.07	0.14	0.09	0.20	0.18
R	0.23	0.38	0.16	0.38	0.45	0.30

* The standardized regression coefficients reported in Table B3 were obtained from the same equations that produced the unstandardized regression coefficients reported in Table 5.

TABLE B4. Standardized Regression Coefficients and Sample Size (N) for Variables Used in Predicting Self-Reported Preventive Health Behavior*

Predictor Variables	Dependent Variables					
	Preventive Health Advice From Doctor (N = 839)	No. of Years Since Physical Exam (N = 839)	Stopped Smoking (N = 490)	Smoking Less (N = 322)	Practice BSE (N = 440)	Practice Monthly BSE (N = 440)
"Mode of payment"	0.01	−0.04	−0.03	−0.03	−0.04	0.00
Age	0.10†	−0.05	0.01	−0.01	−0.03	0.01
Sex	−0.01	−0.15‡	−0.11§	−0.20‡		
Marital status	−0.01	−0.05	−0.06	0.05	0.09	0.04
Socioeconomic status	0.00	−0.04	−0.03	0.10	0.06	0.02
Race/ethnicity	0.06	0.00	0.00	−0.14§	0.01	−0.01
Income	0.01	−0.03	−0.02	−0.05	−0.13§	−0.03
Bed days	0.00	0.04	0.01	0.08	−0.05	−0.07
Reduced activity days	0.05	−0.10†	−0.04	0.07	0.02	0.03
Self-rated health status	0.19‡	0.03	0.11§	0.00	0.00	−0.02
Chronic problems	0.09§	−0.07	0.04	0.02	−0.04	−0.02
Acute problems	0.02	−0.02	−0.01	0.15§	−0.02	−0.05
R	0.30	0.24	0.20	0.31	0.14	0.11

BSE: Breast self-examination.

* The standardized regression coefficients reported in Table B4 were obtained from the same equations that produced the unstandardized regression coefficients reported in Table 6.

† $p < 0.01$.

‡ $p < 0.001$.

§ $p < 0.05$.

MEDICAL CARE
October 1981, Vol. XIX, No. 10

Use of Ambulatory Medical Care by the Poor: Another Look at Equity

JOEL C. KLEINMAN, PH.D.,* MARSHA GOLD, SC.D.,† AND DIANE MAKUC, DR. P.H.‡

Access to health services by the poor and other disadvantaged groups has improved considerably over the past 15 years. These circumstances have led some to question whether the poor now have equal access to health care. This article presents recent evidence from the 1976–78 National Health Interview Surveys (NHIS) comparing utilization among age, race, and income groups. Without adjustment for health status, the poor have more physician visits than those with higher incomes. After adjusting for age and health status, however, these differences are reversed. Depending on which measure is used, the poor have between 7 per cent and 44 per cent fewer visits than those with income above twice the poverty level. Furthermore, the age- and health-adjusted data show blacks have significantly fewer visits than their white counterparts. In addition, there are large differences among race and income groups in the characteristics of the ambulatory care obtained. Blacks and the poor are much more likely to use hospital clinics and less likely to use private physician offices or telephone consultations. The poor also receive less preventive care. It would appear from the present evidence that still further progress is required to achieve the goal of equity in the distribution of medical care services.

ACCESS to health services by the poor and other disadvantaged groups has improved considerably over the past 15 years. These circumstances have led some to question whether the poor now have equal access to health care. In a recent book, Aday, Andersen and Fleming[1] conclude that while some problems of access persist, utilization of medical services is fairly equitable. Considerable controversy has surrounded this statement. Although most observers agree that gaps in access have narrowed during the past 15 years, substantial disagreement exists on the size, nature and importance of remaining problems.[1–5]

The purpose of this article is to present new results on access to medical care by the poor based on the National Health Interview Surveys for 1976-78. We also place these results in the context of previously published work, especially that of

---

* Division of Analysis, National Center for Health Statistics, Department of Health and Human Services.

† Office of Deputy Assistant Secretary for Planning and Evaluation/Health (ODASPE/H), Department of Health and Human Services.

‡ Division of Environmental Epidemiology, National Center for Health Statistics, Department of Health and Human Services. (Dr. Makuc's work on this article was performed while she was with ODASPE/H).

The views expressed in this article are those of the authors, and no official endorsement by the Department of Health and Human Services is intended or should be inferred.

Address for reprints: Joel C. Kleinman, Ph.D., Director, Division of Analysis, National Center for Health Statistics, FCB#3, Room 2-27, 3700 East-West Highway, Hyattsville, MD 20782.

0025-7079/81/1000/1011/$01.45 © J. B. Lippincott Co.

Aday et al.[1] Because the focus is on primary and preventive care, we do not consider use of hospital services.

Data And Methods

Data Sources

The basic data source used is the National Health Interview Survey (NHIS), which is conducted by the National Center for Health Statistics.[6] NHIS is a probability sample of the civilian noninstitutional population with 110,000 respondents in 40,000 households each year. The NHIS data presented in this article were collected during the period 1976 to 1978. Data for all three years were combined in order to minimize sampling error.

Since we will compare our results with those presented by Aday et al.,[1] it is important to note some methodological differences between NHIS and their 1976 survey (carried out by the Center for Health Administration Studies [CHAS] and the National Opinion Research Center [NORC] at the University of Chicago. The CHAS–NORC survey was primarily designed to measure access to medical care. It included 7,787 individuals in 5,432 households but oversampled certain population subgroups believed to have access problems.

The two most salient differences between the surveys are the levels of family income used in analysis and the measurement of physician visits. The NHIS data presented here categorize family income to approximate the official poverty income level adjusted for family size. For example, poverty was defined at $7,000 for a family of four and $4,000 for a family of two. Data published from CHAS–NORC generally use a low-income-category ceiling of $8,000 with no adjustment for family size. The NHIS data result in an estimated 16 per cent of the population below poverty, while the CHAS–NORC low-income group comprises one third of the population. At the other end of the scale, the highest income category used in the NHIS tables (more than 200 per cent of the poverty level) includes 52 per cent of the population, while the CHAS–NORC tables include one third of the population in the highest category.

In defining physician visits, the NHIS data are based on a 2-week recall period while the CHAS–NORC data use a full year recall. Although even the 2-week recall underestimates physician visits, the greater recall bias inherent in the CHAS–NORC method results in a greater underestimate of total physician visits and tends to diminish differentials among population subgroups. The NHIS data include telephone consultations with a physician (excluding calls for making an appointment) while the CHAS–NORC data do not. The inclusion of telephone consultations is particularly important for assessing race and income differentials, since white and high-income persons make more use of the telephone for physician contacts.

The Role of Health Status

Access to health care for any given group can only be assessed relative to the need for such services. Comparisons of utilization among population groups need to consider differentials in health status among those groups. Using a variety of different measures, numerous studies have shown that the poor have worse health status than the nonpoor.[2,3,7]

Health status indicators that measure differences in need for ambulatory medical care are not readily available. Physicians spend much of their time helping people to function more effectively and relieving pain and discomfort. Thus, measures of functional ability and positive well-being are the relevant measures of health status given the objectives of medicine today.

Some promising work in this area has been carried out recently. For example, Brook et al.[8] have developed measures of physical, mental and social health status that were designed to be sensitive to the utilization of medical care. These measures have been found to be reliable and valid and are based on data from self-administered questionnaires. Unfortunately enough data on these measures are not yet available to compare the poor and nonpoor.

Manning, Newhouse and Ware[9] suggest that self-assessment of health is a good predictor of the probability of using health services and of the quality of services demanded, as well as correlating well with clinical indications of need for care and other measures of health status. Since these are the types of measures available from the NHIS, we shall use them in this study.

A further problem in examining income differentials in health status is the income-depressing effects of illness. That is, illness may lead to loss of work, which in turn reduces income. This phenomenon is especially evident in the age groups 17–64. However, it is relatively absent in the under-17 age group.

Adjusting Utilization for Health Status

Because the poor are less healthy than the nonpoor, it is important to adjust for health status in examining utilization data. Two methods of adjustment will be used in this article. The first method examines the use of physician services after stratifying the total population into subgroups that are similar in their overall health status. We will use the respondent's self-assessment of health to stratify the population into two groups. The limitation of this approach is that even within the strata there may be variation in need. Nonetheless, this method provides a useful measure that can be easily interpreted.

The second method of adjusting for health status uses a ratio, which indicates how a group's utilization compares against some measure of need for services by this group. The most commonly used indicator of this type is the "use disability ratio" (UDR).[10,11] The UDR measures the number of physician visits per 100 disability days for those who have had at least one but no more than 100 disability days.[1] A similar measure will be presented from the 1978 NHIS data. This measure is the ratio of physician visits per 100 bed-disability days. Bed-disability days are used rather than all disability days since we believe the former better represent conditions requiring medical attention and are subject to fewer differences in interpretation across income groups. The "use-bed-disability ratio" (UBDR) includes hospital days, which are excluded from the use-disability ratio. Furthermore, it compares visits to bed-disability days for an entire population subgroup, not just those with at least one bed-disability day.

While useful in summarizing overall patterns, use-disability measures suffer from serious limitations.[3,10] The number of disability days represents only a partial measure of need for physician services. In particular, it is relatively insensitive to differences in disease severity of underlying health status contributing to disability. Furthermore, the indicator assumes that disability days and visits are related in a linear fashion. This assumption omits any consideration of visits that are for preventive purposes, or disabilities that do not require medical care.

Statistical Analysis

Three variables were generally used to cross-classify the data: age (under 17, 17–44, 45–64, 65 and older), race (white and black; other races were omitted due to small sample size) and family income (below poverty level, between poverty and 150 per cent of poverty, between 150 per cent and 200 per cent of poverty, above 200 per cent of poverty; respondents who did not report income [9 per cent of total] were omitted).

When the data are cross-classified in this detail, many of the individual cell estimates are subject to considerable sampling error. As a result, the values observed in certain cells may not be good estimates of the corresponding true values. A powerful statistical technique for "smoothing" the cell estimates is the generalized linear model fitted by a weighted least-squares algorithm.[12-14] These smoothed estimates "immediately identify for the substantive analyst statistically significant differences while suppressing nonsignificant or purely sampling differences."[14] The results reported in the next section are based on the smoothed estimates from these models. Details of the model fitting procedure and observed (i.e., unsmoothed) data are given in the appendix.

In order to express differences in relative terms (i.e., per cent differences), logarithms of the response measures in each cell (e.g., mean number of doctor visits) were fit to a full linear model. Interaction terms were omitted from the model in a stepwise procedure until only effects that were statistically significant (at the 10 per cent level) remained.

Results

Self-Assessment of Health

NHIS respondents were asked (for themselves and other household members) "compared to other persons ______'s age, would you say that his health is excellent, good, fair or poor?" The estimated proportion reporting fair or poor health by age, race and income is shown in Table 1. The general patterns by age are similar: fewer whites than blacks report fair or poor health. There is also a steady decrease in the proportion reporting fair or poor health as income increases. The effect of income on reported health status is greater for whites than blacks.

Among whites the ratios of the proportion reporting fair or poor health in the lowest to that in the highest income group vary from 2.1 for those 65 and over to 3.7 for the 45–64 age group. For blacks the ratios range from 1.5 for those 65 and over to 2.6 for adults 45–64 years of age. For both races, the largest ratios occur in the two age groups between 17 and 64. This is consistent with the income-depressing effects of illness noted earlier. The two youngest age groups show the largest differences by race.

Other measures of health, such as proportion with limitation of activity due to chronic illness, restricted activity and bed-disability days, exhibit similar patterns. Hospitalization rates and length of hospital stays are also greater for blacks and low-income persons, providing further evidence of more serious illness in these groups.

Physician Visits

The NHIS indicates that without adjustment for age or health status, physician visits for the poor slightly exceed those of higher income people. Those below poverty in 1976-78 made 5.1 physician visits per year while those with incomes over twice the poverty level made 4.9 physician visits.

These overall figures mask significant variations by age and race (Table 2). The aggregate differences in use reflect below-average use by poor children balanced against the higher utilization for poor nonelderly adults. For children under 17, visits increase with income and are 23 per cent lower for blacks than whites. Within each race, those in the two lowest income groups have 26 per cent fewer visits than those in the highest category.

For the 17–44 age group, blacks have 17 per cent more visits than whites. Furthermore, the income differential reverses, with the poor having 37 per cent more visits than the two highest income groups.

TABLE 1. Estimated* Per Cent Reporting Fair or Poor Health by Age, Race and Income, 1976–78

Race and Income	Age Adjusted†	Under 17	17–44	45–64	65 and Older
White					
Below poverty	23.5	7.2	16.9	49.5	42.2
100–150% poverty	16.5	4.6	10.7	35.9	33.3
150–200% poverty	12.6	3.2	8.2	26.6	27.4
Above 200% poverty	7.5	2.5	4.8	13.4	19.9
Black					
Below poverty	28.4	9.6	22.9	55.0	48.7
100–150% poverty	22.6	6.9	16.5	45.5	43.9
150–200% poverty	18.4	5.2	13.5	36.1	38.7
Above 200% poverty	12.8	4.7	9.2	21.1	32.8

* Estimated from generalized log-linear model.

† Age-adjusted by the direct method.

In the two oldest age groups there is no evidence of a race differential. For persons 45–64 years of age, visits decrease steadily with income, with the poor having 42 per cent more visits than the highest income group. For the population 65 and over there are no income differences in utilization.

Thus, poor adults have at least as many physician visits as the nonpoor. Children in families with incomes below 150 per cent of the poverty level, however, have fewer visits than those in higher-income families. These aggregate visit data do not take into account the lower health status among the poor noted in the previous section.

The volume of visits reflects both the proportion of the population that visits a physician and the number of visits made by persons seeing physicians. The first component is a measure of access to the health care system which focuses on patient-initiated visits, while volume includes a substantial proportion of physician-initiated visits for follow-up care. Although

TABLE 2. Estimated* Number of Physician Visits per Person per Year by Age, Race and Income, 1976–78

Race and Income	Age Adjusted†	Under 17	17–44	45–64	65 and Older
White					
Below poverty	5.39	3.66	5.24	7.40	6.69
100–150% poverty	4.95	3.66	4.75	6.25	6.69
150–200% poverty	4.82	4.06	4.51	5.54	6.69
Above 200% poverty	5.00	4.92	4.51	5.22	6.69
Black					
Below poverty	5.37	2.81	5.78	7.40	6.69
100–150% poverty	4.91	2.81	5.24	6.25	6.69
150–200% poverty	4.74	3.12	4.97	5.54	6.69
Above 200% poverty	4.87	3.78	4.97	5.22	6.69

* Estimated from generalized log-linear model.

† Age-adjusted by the direct method.

TABLE 3. Estimated* Per Cent With No Physician Visits in the 2 Years Preceding Interview by Age, Race and Income, 1976–78

Race and Income	Age Adjusted†	Under 17	17–44	45–64	65 and Older
White					
Below poverty	15.6	14.6	14.7	19.2	14.9
100–150% poverty	15.2	13.5	14.7	19.2	14.2
150–200% poverty	13.6	10.5	13.7	17.6	13.4
Above 200% poverty	11.1	7.0	12.2	14.4	11.7
Black					
Below poverty	13.6	14.6	11.9	14.8	14.9
100–150% poverty	13.2	13.5	11.9	14.8	14.2
150–200% poverty	11.7	10.5	11.2	13.5	13.4
Above 200% poverty	9.5	7.0	9.9	11.1	11.7

* Estimated from generalized log-linear model.

† Age-adjusted by the direct method.

both are important components of care, the proportion of the population with no visit may be more affected by barriers to access.

In order to identify relatively serious access problems, we examine the proportion who have not seen a physician within the 2 years preceding interview (Table 3). This proportion is greater among those below the poverty level compared with those above twice the poverty level. The income trend shows in all age groups, especially among whites, but the differential is greatest among children. They are twice as likely as their high-income counterparts to have had no visit in the past 2 years. There are no race differences among children or the elderly, but the proportion with no visit is greater among whites than blacks in the 17–64 age groups.

Physician Visits Stratified by Health Status

Table 4 shows the estimated number of physician visits for respondents reported in fair or poor health. For children under 17 there is a steady increase in visits for both blacks and whites—poor children have 47 per cent fewer visits than children in high-income families. Furthermore, at each in-

TABLE 4. Estimated* Number of Physician Visits per Person per Year for Those Reporting Fair or Poor Health by Age, Race and Income, 1976–78

Race and Income	Age Adjusted†	Under 17	17–44	45–64	65 and Older
White					
Below poverty	9.98	9.14	10.36	10.95	8.90
100–150% poverty	10.68	11.36	10.36	10.95	9.61
150–200% poverty	11.34	13.30	10.36	10.95	10.66
Above 200% poverty	13.51	17.16	12.69	10.95	11.90
Black					
Below poverty	8.97	5.68	11.39	9.39	7.58
100–150% poverty	9.42	7.06	11.39	9.39	8.18
150–200% poverty	9.85	8.26	11.39	9.39	9.08
Above 200% poverty	10.65	10.67	11.39	9.39	10.14

* Estimated from generalized log-linear model.

† Age-adjusted by the direct method.

come level, blacks have 38 per cent fewer visits than whites.

Among blacks age 17–44 who report fair or poor health, there is no statistically significant income differential. For whites in this age group the three lowest income groups have the same number of visits but the highest income group has 23 per cent more visits. The number of visits for blacks falls midway between the low- and high-income whites.

For the age group 45–64, blacks experience 14 per cent fewer visits than whites and there are no income differentials for either race. The elderly experience an increasing number of visits as income increases, with the poor having a 25 per cent deficit relative to the highest income group. Blacks make 15 per cent fewer visits than whites (although this effect is just short of being statistically significant at the 5 per cent level).

The smoothed utilization patterns for those reporting good or excellent health (Table 5) are different in certain respects from the patterns for those reporting fair or poor health. For children there are similarities in that blacks experience fewer visits than whites (by 24 per cent) and the income effect is large (the poor have 29 per cent fewer visits than the nonpoor). The major differences are that the two lowest income groups exhibit the same number of visits and the remaining differences are smaller than those for children in fair or poor health.

For the 17–44 and 45–64 age groups there are no significant differences by race or income. For the elderly there are no differences by income but blacks have 18 per cent more visits than whites.

To summarize, the data show that, before adjustment for health status, blacks and the poor aged 17–64 use more physician visits than their counterparts. The effect of stratifying by self-assessed health status is to reverse or at least eliminate these differences. For children even the unadjusted data showed large income and racial differentials, but these gaps increase for those reporting fair or poor health. Furthermore, the elderly who report fair or poor health exhibit a steady increase in visits with increasing income, and blacks in this group also appear to have fewer visits. Although the stratification of the population into two health status categories (fair or poor versus good or excellent) does not result in groups with equal need for medical care, even such a crude adjustment suggests income and racial differences that are not apparent in the unadjusted data.

TABLE 5. Estimated* Number of Physician Visits per Person per Year for Those Reporting Good or Excellent Health by Age, Race and Income, 1976–78

Race and Income	Age Adjusted†	Under 17	17–44	45–64	65 and Older
White					
Below poverty	3.99	3.27	4.09	4.16	5.21
100–150% poverty	3.99	3.27	4.09	4.16	5.21
150–200% poverty	4.13	3.75	4.09	4.16	5.21
Above 200% poverty	4.37	4.60	4.09	4.16	5.21
Black					
Below poverty	3.87	2.49	4.09	4.16	6.17
100–150% poverty	3.87	2.49	4.09	4.16	6.17
150–200% poverty	3.98	2.86	4.09	4.16	6.17
Above 200% poverty	4.16	3.51	4.09	4.16	6.17

* Estimated from generalized log-linear model.

† Age-adjusted by the direct method.

Use–Bed Disability Ratio

The age-adjusted number of physician visits per 100 bed-disability days (UBDR) is 80 per cent higher among high-income whites than among those below the poverty level. For blacks, the differential is about 40 per cent. However, there are significant differences by age and race. Among children, whites have a higher UBDR than blacks in all income groups except those below poverty. Furthermore white children show a sharp increase in UBDR as income increases: the UBDR for the highest income group is nearly twice that for the poor. For black children, however, there is no evidence of an income differential.

In the 17–44 age group the UBDR for whites is 26 per cent higher than that for blacks in each income group. Both whites and blacks show a strong income effect: the UBDR for the high income group is 56 per cent higher than that for the poor.

White persons aged 45–64 generally have a higher UBDR than blacks, although the difference varies with income. Whites show a very strong increase in UBDR with increasing income (the ratio of the UBDR for the highest to lowest income groups is 2.5). The income effect for blacks appears to be limited to a deficit of 34 per cent among the poor, with no difference among the three higher income groups.

Blacks and whites in the oldest age group also have different trends by income. Among whites the UBDR shows a steady increase with a high-to-low ratio of 1.76. For blacks, however, there appears to be no difference within the two highest and two lowest income groups. The ratio between these high and low values is 2.18.

It may be useful to note at this point that these results differ from the CHAS–NORC data, which show no income differentials in the use-disability ratio.[1] The reasons for this discrepancy are probably related to the different survey methodologies previously discussed: the greater recall bias and omission of telephone contacts in CHAS–NORC and the use of broader income classes with no adjustment for family size by Aday et al.[1]

Sources of Care

The quantity of physician services used reflects only a part of the concern over access to care. As David Rogers remarked in his testimony to the Subcommittee on Health and Scientific Research of the United States Senate:

> Simply counting the number of visits people make to doctors each year is not

TABLE 6. Estimated* Number of Physician Visits per 100 Bed-Disability Days by Age, Race and Income, 1976–78

Race and Income	Age Adjusted†	Under 17	17–44	45–64	65 and Older
White					
Below poverty	53.0	53.5	64.5	38.4	34.9
100–150% poverty	63.5	69.9	73.8	44.2	44.2
150–200% poverty	73.9	82.1	85.6	51.2	50.8
Above 200% poverty	95.1	99.8	100.7	94.5	61.5
Black					
Below poverty	48.5	63.3	51.3	32.2	29.4
100–150% poverty	54.9	63.3	58.7	48.6	29.4
150–200% poverty	62.3	63.3	68.0	48.6	64.0
About 200% poverty	67.2	63.3	80.1	48.6	64.0

* Estimated from generalized log-linear model.

† Age-adjusted by the direct method.

> sufficient. The term "physician visit" can mean many things. It can be a hurried visit to a "Medicaid mill" in which a patient is ping-ponged among a series of doctors, nurses and others to maximize reimbursement income. It can be hours spent waiting in a crowded, noisy hospital outpatient clinic or a few minutes with a doctor you've never seen before. At the other end of the scale–and far more desirable personally and medically–is the kind of visit you and I are accustomed to: seeing a doctor we know by name and who knows us and our families. Implicit in this relationship is the physician's assumption of personal responsibility for attending to an individual's medical care needs.[15]

The aggregate visit data discussed earlier encompass wide variation by source of care. Table 7 shows certain characteristics of visits by race and income (observed data are shown since the large and systematic differences in the observed data obviated the need to use the smoothing techniques discussed previously). Although the majority of visits take place in an office setting regardless of race or income, blacks are twice as likely as whites to use hospital outpatient departments or emergency rooms and half as likely to use the telephone. Furthermore, within each race, there is a consistent trend by income. These differentials are particularly large for children, especially those under age 6. In this group 25 per cent of the visits for whites above twice the poverty level are by telephone, compared with 14 per cent of those below the poverty level. The corresponding proportions for blacks are 20 per cent and 4 per cent. The range in proportion of visits at hospital clinics was 7 per cent (high income) to 18 per cent (poverty) for whites and 16 per cent to 36 per cent for blacks.

Conclusions

The NHIS data for 1976-78 indicate that income and race continue to influence the use of ambulatory medical care. Without adjusting for health status, the 1976-78 NHIS shows that the poor have more physician visits than those with higher incomes. After adjusting for age and health status, however, these differences are reversed. Depending on which measure is used, the poor have between 7 per cent and 44 per cent fewer visits than those with income above twice poverty level. Furthermore, the age- and health-adjusted data show that blacks have significantly fewer visits than their white counterparts.

For children under 17, even the unadjusted data show a sizeable deficit in

TABLE 7. Characteristics of Physician Visits by Race and Income, 1976–78

Race and Income	Physician's Office (% of visits)	Telephone (% of visits)	Hospital Clinic or Emergency Room (% of visits)
White			
Below poverty	64.1	10.8	15.3
100–150% poverty	67.2	11.1	14.0
150–200% poverty	68.7	12.9	13.3
Above 200% poverty	71.0	13.6	9.8
Black			
Below poverty	52.7	3.6	31.5
100–150% poverty	56.4	5.8	26.7
150–200% poverty	60.1	6.6	22.8
Above 200% poverty	64.8	7.9	18.7

physician visits among blacks and the poor. After adjusting for health status the income and racial differentials become larger. The unadjusted data for the elderly show no income or race effects. For the elderly who report fair or poor health, however, deficits among the poor and blacks become evident.

The methods used here to adjust utilization for health status are admittedly rather crude. Nonetheless, compared with the unadjusted data, they resulted in a different perspective on whether the quantity of ambulatory care received by the poor is commensurate with their need. Furthermore, these conclusions are consistent with the results of analyses using other methods to adjust utilization data through multiple regression and its econometric variants. These methods are generally used to simultaneously adjust for a large number of variables, including health status. Several studies using these methods have also shown that utilization of physician services by the poor lags substantially behind that of the nonpoor.[2]

Despite the plethora of studies on the measurement of use relative to need, the state of the art is still in flux. No single method or measure has achieved universal acceptance. The difference in results between the apparently similar use-disability ratio (CHAS–NORC, 1976) and use-bed disability ratio (NHIS, 1976–78) underscores the danger in relying on a single methodology in addressing this issue.

One set of differences between the poor and nonpoor that holds up consistently across many different studies and over several years are those involving characteristics of care. In all age groups there are substantial race and income differences in the type of visits. According to the 1976-78 NHIS data, blacks and the poor are more likely to use hospital clinics and less likely to use private physician offices or telephone consultations. The 1976 CHAS–NORC survey documented substantial deficits by the poor in access to regular source of care and satisfaction with medical care.[1,2]

The second consistent result is a deficit by the poor in use of preventive procedures. Several studies have shown that, compared to those with higher incomes, the poor are less likely to receive breast exams, Pap tests, prenatal care or immunization against childhood diseases (particularly in the preschool years).[2] Dental care is another area where large differences remain and for which little progress has been made over the past 15 years.[2,16,17]

These indications that substantial disparities in utilization and access to care still exist for low-income as well as other disadvantaged groups raise significant policy concerns. Aday et al.[1] suggest the following definition: "an equitable distribution of health care services is one in which illness is the major determinant of the allocation of resources." This does not imply that medical care alone can cure all illness or alleviate all pain. Nutrition, environmental exposures, housing, personal habits, education and other factors influence health status. Even if medical care were allocated in an equitable fashion, equal health outcomes for the poor and nonpoor could not be assured.

Nevertheless, medical care can contribute to the reduction of certain types of mortality, morbidity and functional impairment. Medical care can also alleviate pain and anxiety, perhaps explaining why public opinion polls have shown that a large proportion of the U.S. population places a high value on medical care. It would appear from the present evidence that still further progress is required to achieve the goal of equity in the distribution of medical care services.

Acknowledgments

The authors would like to thank Karen Davis and Jack Feldman for their advice, support and critical review throughout the course of this project. Bruce Cohen, Margaret Cooke and Ron Wilson provided valuable comments on earlier drafts of this article.

Bonnie Brickett, Mitchell Pierre, Jr., Patricia Szafranski and Jean Williams provided data processing and research assistance for this study.

References

1. Aday LA, Andersen R, Fleming GV. Health care in the U.S.: equitable for whom? Beverly Hills, Calif.: Sage Publications, 1980.

2. Davis K, Gold M, Makuc D. Access to health care for the poor: does the gap remain? Ann Rev Public Health 1981;2:159.

3. Davis K, Schoen C. Health and the war on poverty: a ten year appraisal. Washington, D.C.: The Brookings Institution, 1978.

4. Somers HM. Health and public policy. Inquiry 1975;12:87.

5. Fuchs V. Who shall live? New York: Basic Books, 1974.

6. National Center for Health Statistics. Current estimates from the Health Interview Survey, 1978. Vital and health statistics, Series 10, no. 130. DHEW publication no. (PHS) 80-1551, November 1979.

7. Kitagawa EM, Hauser PM. Differential mortality in the United States: a study in socioeconomic epidemiology. Cambridge, Mass.: Harvard University Press, 1973.

8. Brook RH, Ware JE, Davis-Avery A et al. Conceptualization and measurement of health for adults in the health insurance study: Vol. VIII, overview. Santa Monica, Calif.: Rand Corporation, 1979. (DHEW publication no. R-1987/8-HEW).

9. Manning WG, Newhouse JP, Ware JE. The status of health in demand estimation: beyond excellent, good, fair, and poor. In: Fuchs VR, ed. Economic aspects of health. Chicago: University of Chicago Press, 1931.

10. Andersen R. Health status indices and access to medical care. Am J Public Health 1978;68:458.

11. Aday LA, Andersen R. Development of indices of access to medical care. Ann Arbor: Health Administration Press, 1975.

12. Koch GG, Freeman DH, Freeman JL. Strategies in the multivariate analysis of data from complex surveys. International Statistical Review 1975;43:59.

13. Freeman DH, et al. Strategies in the multivariate analysis of data from complex surveys II: an application to the U.S. National Health Interview Survey. International Statistical Review 1976;44:317.

14. Freeman DH, et al. A linear model methodology for the analysis of physician visit data from NCHS. Am J Pub Health 1976;66:979.

15. Rogers DE. Testimony before the Subcommittee on Health and Scientific Research, Committee on Labor and Human Resources, U.S. Senate, 1980.

16. Wilson RW, White EJ. Changes in morbidity, disability, and utilization differentials between the poor and nonpoor. Med Care 1977;15:636.

17. Madans J, Kleinman JC. Utilization of ambulatory medical and dental services among the poor and nonpoor. Health, United States, 1980. Washington, D.C.: U.S. Department of Health and Human Services. (DHHS publication no. (PHS) 81-1232, October 1980.)

Appendix

In this section we present the details of the model-fitting procedures used to obtain the smoothed estimates given in the body of the article. A complete discussion of the rationale for applying linear model techniques to survey data can be found in references 12–14. Freeman et al.[13] point out the need to use appropriate estimates of the complete covariance matrix in such analyses. Unfortunately the computations involved in estimating such matrices were beyond the resources available for this study. Instead we use approximate variance curves derived for HIS[6] and assume zero covariance between cell estimates. Freeman et al.[13] found that the zero-covariance assumption tends to result in a loss of power. Thus, the test statistics we use are generally conservative and we run the risk of omitting significant effects. On the other hand, the multiple significance tests being carried out on the data may result in detecting too many "significant" differences.

We use a backward elimination procedure in model fitting by starting with a fully saturated model and reducing interaction terms that are not significant (at the 10 per cent level) in a hierarchical manner (i.e., three-way interactions, then two-way, etc.). If interactions with age were present we fit models separately within age group. Although this is not a standard practice in formal statistical modeling, we felt the nature of the measures used reflected substantive differences by age (e.g., the type of poor health reported for a child may differ from that reported for an elderly person). If fitted income effects did not follow a monotonic pattern, we tested adjacent categories to see if they could be combined.

After eliminating all nonsignificant terms we examined the table of residuals to look for unusual patterns or outliers. Since the variances of estimates for blacks were substantially larger than those for whites, residuals for blacks were generally larger. If the examination of residuals suggested a pattern, terms were added to the model to improve its fit. While this approach invalidates a strict statistical-significance-level interpretation of the effects, we feel it provides a reasonable compromise between formal modeling and minimizing the risk of omitting real differences.

Table A1 shows the observed means or proportions for each cell in the age × race × income tables and Table A2 shows the corresponding standard errors. The results of the model-fitting procedures for each variable are outlined in tables A3–A7 and discussed in the following text. The abbreviations used in the tables are as follows:

P = below poverty level
1.5P = 100—150 per cent poverty
2P = 150—200 per cent poverty
NP = more than 200 per cent poverty
POP = population (in thousands)
%FP = per cent reporting fair or poor health
MDV = number of physician visits
% no MDV = per cent with no physician visits in 2 years preceding interview
MDV/FP = number of physician visits for those reporting fair or poor health
MDV/GE = number of physician visits for those reporting good or excellent health
MDV/BD = number of physician visits per 100 bed-disability days
A = age
R = race
I = income

Self-Assessed Health

Table A3 shows the results of model fitting for the proportion reporting fair or poor health. After eliminating the three-way interaction terms, each of the two-way interactions were highly significant. Since the race × income interaction terms within each age were also significant, we did not attempt to reduce the model with no three-way interaction any further. The residuals were all below 5 per cent for whites and 14 per cent for blacks.

Physician Visits

Main effects models within each age group provided an adequate fit to the data on number of physician visits (Table A4). For children under 17 and adults age 17–44, two adjacent income categories could be combined. Residuals were all below 3 per cent for whites and, except for two cases, below 10 per cent for blacks.

For the per cent with no physician visit in the past two years (Table A5), the combined race × income effect, given no three-way interaction, was significant at the 0.066 level. However, there was no evidence of a race × income interaction within any age group, so we proceeded to fit separate main effects models within each age. The residuals were below 3 per cent for whites. Among blacks there were three residuals between 15 per cent and 20 per cent, but all the rest were below 10 per cent.

For those reporting fair or poor health, the model with $ARI = RI = 0$ fit well (Table A6). However, when main-effects models were fit separately within each group, there was a significant race × income interaction for those 17–44. This interaction resulted from a positive relationship between visits and income for whites but a negative relationship for blacks in the 17–44 age group. When these income effects are examined separately, the effect for whites is significant ($p = 0.002$), while that

TABLE A1. Selected Health Characteristics by Age, Race and Income, 1976–78

	POP	% FP	MDV	% No MDV	MDV/FP	MDV/GE	MDV/BD
Both races							
All ages							
P	29560.36	22.06	5.07	14.71	9.75	3.74	46.82
1.5P	23311.58	17.73	4.95	14.74	10.19	3.82	54.80
2P	38203.75	12.30	4.73	13.08	10.37	3.94	67.28
NP	98763.69	7.59	4.93	11.40	12.09	4.34	92.31
Under 17							
P	11237.52	8.10	3.42	14.61	7.66	3.05	55.44
1.5P	7222.80	4.96	3.40	13.51	10.18	3.04	66.63
2P	12390.35	3.43	3.97	10.51	12.59	3.66	81.68
NP	23023.05	2.61	4.85	6.99	16.50	4.53	98.09
17–44							
P	10561.60	18.56	[illegible]	13.66	10.98	4.12	59.73
1.5P	8131.35	11.55	[illegible]	14.54	11.02	4.01	70.71
2P	15187.08	8.71	4.46	13.45	10.27	3.90	83.16
NP	45042.41	5.09	4.57	12.05	12.26	4.16	99.12
45–64							
P	3680.25	50.98	7.39	17.88	10.54	4.12	36.62
1.5P	3374.80	37.52	6.25	18.55	10.51	3.69	45.44
2P	6033.45	27.58	5.53	17.21	9.77	3.92	49.64
NP	24984.88	13.74	5.22	14.22	11.29	4.25	80.91
65 and over							
P	4080.99	43.46	6.66	14.87	8.62	5.15	33.57
1.5P	4582.63	34.25	6.65	14.24	9.45	5.19	42.49
2P	4592.87	28.00	6.62	13.36	10.52	5.10	50.98
NP	5713.43	20.35	6.80	11.71	11.80	5.52	62.26
White							
All ages							
P	20759.48	21.76	5.21	15.39	9.93	3.90	47.67
1.5P	19346.57	17.15	4.99	14.83	10.33	3.88	55.58
2P	34548.60	11.88	4.73	13.23	10.48	3.95	68.04
NP	92787.44	7.34	4.95	11.55	12.36	4.37	94.44
Under 17							
P	7188.61	7.30	3.74	15.00	8.86	3.33	53.53
1.5P	5962.82	4.33	3.56	13.30	12.00	3.17	69.85
2P	11185.88	3.27	4.02	10.46	13.24	3.71	82.07
NP	21569.42	2.50	4.94	7.08	17.19	4.62	99.76
17–44							
P	7600.13	16.73	5.15	14.52	10.18	4.14	64.00
1.5P	6890.39	10.56	4.69	14.67	10.73	3.98	72.04
2P	13612.24	8.22	4.41	13.73	10.31	3.88	84.71
NP	42009.13	4.83	4.57	12.22	12.69	4.16	102.03
45–64							
P	2674.02	49.62	7.63	19.13	11.18	4.13	38.35
1.5P	2815.03	36.34	6.21	19.29	10.46	3.79	44.19
2P	5416.59	26.61	5.54	17.50	9.82	3.99	51.18
NP	23692.88	13.25	5.23	14.41	11.51	4.27	94.47
65 and over							
P	3296.72	42.27	6.59	15.20	8.91	4.89	34.88
1.5P	4178.33	33.37	6.70	14.27	9.73	5.18	44.19
2P	4333.89	27.19	6.53	13.43	10.58	5.02	50.79
NP	5516.08	19.97	6.76	11.65	11.82	5.50	61.46
Black							
All ages							
P	8800.87	22.77	4.74	13.13	9.34	3.38	44.75
1.5P	3465.01	21.07	4.71	14.25	9.54	3.43	50.52

TABLE A1. Continued

	POP	% FP	MDV	% No MDV	MDV/FP	MDV/GE	MDV/BD
2P	3655.15	16.24	4.74	11.70	9.65	3.79	60.91
NP	5976.26	11.41	4.56	9.00	9.32	3.94	66.85
Under 17							
P	4048.92	9.52	2.87	13.93	6.03	2.54	60.42
1.5P	1259.98	7.92	2.64	14.46	5.45	2.40	51.49
2P	1204.47	4.89	3.46	10.98	8.61	3.20	77.72
NP	1453.64	4.24	3.46	5.73	10.44	3.15	72.40
17–44							
P	2961.47	23.27	6.02	11.44	12.45	4.07	52.11
1.5P	1240.96	17.06	5.56	13.87	12.02	4.23	65.08
2P	1574.84	12.98	4.88	11.05	10.07	4.11	72.75
NP	3033.28	8.74	4.65	9.57	8.96	4.24	71.42
45–64							
P	1006.22	54.60	6.76	14.54	8.99	4.07	32.24
1.5P	559.77	43.49	6.44	14.84	10.73	3.14	52.70
2P	616.86	36.10	5.46	14.59	9.41	3.23	39.14
NP	1292.00	22.73	5.06	10.65	8.95	3.92	53.10
65 and over							
P	784.27	48.49	6.92	13.50	7.54	6.35	29.10
1.5P	404.30	43.29	6.20	13.99	7.24	5.40	29.72
2P	258.98	41.48	8.10	12.11	9.92	6.80	53.73
NP	197.35	31.05	7.83	13.42	11.57	6.14	90.42

TABLE A2. Standard Errors of Estimates in Table A1

	POP	% FP	MDV	% No MDV	MDV/FP	MDV/GE	MDV/BD
Both races							
All ages							
P	109.65	0.25	0.08	0.21	0.24	0.07	1.18
1.5P	113.26	0.26	0.09	0.24	0.31	0.08	1.61
2P	89.16	0.18	0.07	0.18	0.30	0.06	1.74
NP	0.0	0.09	0.04	0.11	0.26	0.04	1.31
Under 17							
P	96.45	0.27	0.10	0.35	0.56	0.10	2.67
1.5P	81.52	0.27	0.12	0.42	1.06	0.12	4.23
2P	99.64	0.17	0.10	0.29	1.12	0.10	4.00
NP	113.23	0.11	0.08	0.18	1.13	0.08	3.50
17–44							
P	94.37	0.39	0.13	0.35	0.48	0.13	2.46
1.5p	85.50	0.37	0.14	0.41	0.69	0.14	3.66
2P	105.80	0.24	0.10	0.29	0.55	0.10	3.52
NP	47.89	0.11	0.06	0.16	0.47	0.06	2.73
45–64							
P	60.71	0.86	0.27	0.66	0.47	0.27	1.84
1.5P	58.34	0.87	0.26	0.70	0.58	0.24	2.75
2P	75.60	0.60	0.18	0.51	0.48	0.17	2.48
NP	113.08	0.23	0.09	0.23	0.37	0.08	2.95
65 and over							
P	63.63	0.81	0.24	0.58	0.43	0.28	1.65
1.5P	67.04	0.73	0.23	0.54	0.48	0.24	2.11
2P	67.10	0.69	0.23	0.52	0.37	0.23	2.68
NP	73.85	0.56	0.21	0.44	0.65	0.20	3.10

TABLE A2. Continued

	POP	% FP	MDV	% No MDV	MDV/FP	MDV/GE	MDV/BD
White							
All ages							
P	112.43	0.30	0.09	0.26	0.29	0.09	1.38
1.5P	111.81	0.28	0.09	0.26	0.35	0.09	1.75
2P	100.45	0.18	0.07	0.19	0.32	0.07	1.84
NP	0.0	0.09	0.04	0.11	0.28	0.04	1.39
Under 17							
P	81.36	0.32	0.13	0.44	0.80	0.13	3.03
1.5P	75.22	0.27	0.14	0.46	1.39	0.13	4.84
2P	96.29	0.18	0.11	0.30	1.24	0.11	4.19
NP	112.84	0.11	0.09	0.18	1.22	0.09	3.66
17–44							
P	83.22	0.45	0.15	0.42	0.56	0.15	3.21
1.5P	79.95	0.39	0.15	0.44	0.77	0.15	4.12
2P	102.59	0.25	0.10	0.31	0.60	0.10	3.82
NP	71.19	0.11	0.06	0.17	0.51	0.06	2.92
45–64							
P	52.34	1.01	0.32	0.79	0.58	0.32	2.24
1.5P	53.62	0.94	0.28	0.78	0.64	0.26	2.90
2P	72.17	0.63	0.19	0.54	0.52	0.18	2.71
NP	113.27	0.23	0.09	0.24	0.39	0.08	3.18
65 and over							
P	57.71	0.90	0.27	0.65	0.49	0.29	1.93
1.5P	64.31	0.76	0.24	0.56	0.52	0.25	2.32
2P	65.38	0.70	0.23	0.54	0.60	0.23	2.76
NP	72.74	0.56	0.21	0.45	0.67	0.21	3.11
Black							
All ages							
P	85.19	0.47	0.13	0.38	0.42	0.13	1.95
1.5P	59.05	0.72	0.21	0.62	0.71	0.20	3.57
2P	60.52	0.64	0.21	0.55	0.80	0.20	4.45
NP	75.30	0.43	0.16	0.39	0.73	0.16	4.04
Under 17							
P	63.41	0.48	0.15	0.57	0.74	0.15	5.35
1.5P	36.49	0.79	0.26	1.03	1.37	0.25	8.03
2P	35.70	0.65	0.30	0.94	2.35	0.30	12.48
NP	39.12	0.35	0.28	0.64	2.59	0.27	10.33
17–44							
P	54.90	0.81	0.27	0.61	0.87	0.24	3.56
1.5P	36.23	1.11	0.39	1.02	1.53	0.37	7.64
2P	40.66	0.88	0.32	0.82	1.39	0.31	8.42
NP	55.52	0.53	0.23	0.56	1.13	0.23	6.09
45–64							
P	32.70	1.64	0.49	1.16	0.79	0.54	3.08
1.5P	24.51	2.18	0.64	1.57	1.33	0.56	7.96
2P	25.71	2.02	0.55	1.48	1.28	0.51	5.58
Np	36.94	1.21	0.36	0.89	1.08	0.36	5.98
65 and over							
P	28.94	1.86	0.36	1.27	0.85	0.74	3.03
1.5P	20.86	2.57	0.73	1.80	1.22	0.90	4.55
2P	16.72	3.19	1.08	2.11	1.90	1.27	10.69
NP	14.61	3.43	1.21	2.53	2.78	1.26	25.20

TABLE A3. Summary of Model-Fitting Procedure: Per Cent Reporting Fair or Poor Health

Model	df	χ^2	p
Saturated model			
ARI = 0*	9	10.09	0.343
Model with ARI = 0† (Final model)	9	10.09	0.343
AR = 0*	3	31.12	<0.001
AI = 0*	9	248.65	<0.001
RI = 0*	3	69.85	<0.001
Main effects models for each age:			
Under 17†	3	7.73	0.052
17–44†	3	12.05	0.007
45–64†	3	46.44	<0.001
65 and over†	3	13.73	0.003

* Hypothesis test given specified model.

† These lines give the goodness-of-fit χ^2 for the specified model.

for blacks is not (p = 0.102). There were, however, no significant differences among the three lowest income groups for whites (p = 0.844). Overall, residuals for whites were below 5 per cent except for one cell with a residual of 10 per cent. There were two large residuals (25 per cent) among blacks but both fitted values were within

TABLE A4. Summary of Model-Fitting Procedure: Number of Physician Visits per Year

Model	df	χ^2	p
Saturated model			
ARI = 0*	9	8.12	0.522
Model with ARI = 0†	9	8.12	0.522
AR = 0*	3	57.48	<0.001
AI = 0*	9	198.95	<0.001
RI = 0*	3	3.65	0.302
Main effects model for each age:			
Under 17†	3	2.93	0.403
R = 0*	1	44.42	<0.001
I = 0*	3	109.41	<0.001
P = 1.5P*	1	1.39	0.238
17–44†	3	4.55	0.208
R = 0*	1	11.06	<0.001
I = 0*	3	31.13	<0.001
2P = NP*	1	1.30	0.254
45–64†	3	1.49	0.684
R = 0*	1	0.93	0.335
I = 0*	3	83.01	<0.001
65 and over†	3	2.80	0.424
R = 0*	1	1.18	0.277
I = 0*	3	9.61	0.895
Final model†	19	16.98	0.591

* Hypothesis test given specified model.

† These lines give the goodness-of-fit χ^2 for the specified model.

TABLE A5. Summary of Model-Fitting Procedure: Per Cent With No Physician Visits in Past 2 Years

Model†	df	χ^2	p
Saturated model			
ARI = 0*	9	5.53	0.786
Model with ARI = 0†	9	5.53	0.786
AR = 0*	3	17.16	<0.001
AI = 0*	9	191.79	<0.001
RI = 0*	3	7.18	0.066
Main effects model for each age:			
Under 17†	3	6.08	0.108
R = 0*	1	0.92	0.337
I = 0*	3	487.62	<0.001
17–44†	3	4.25	0.235
R = 0*	1	36.21	<0.001
I = 0*	1	68.26	<0.001
P = 1.5P*	1	1.07	0.301
45–64†	3	0.81	0.848
R = 0*	1	29.40	<0.001
I = 0*	3	97.58	<0.001
P = 1.5P*	1	0.04	0.841
65 and over†	3	1.56	0.668
R = 0*	1	0.65	0.420
I = 0*	3	23.06	<0.001
Final model†	16	15.39	0.496

* Hypothesis test given specified model.

† These lines give the goodness-of-fit χ^2 for the specified model.

the 95 per cent confidence interval for the cell estimate. The remaining residuals were below 10 per cent.

For those reporting good or excellent health, the main effects models within each age group fit very well (Table A7). With only two exceptions, residuals among both whites and blacks were below 13 per cent. The two exceptions were the middle-income groups among blacks aged 45–64 which had residuals of 25 per cent. The fitted value for both cells were within the 95 per cent confidence interval of their corresponding estimate.

Use-Bed-Disability Ratio

The three-way interaction was highly significant (Table A8) so we proceeded to fit separate models within each age group. For children, there was a significant race × income interaction, with whites showing strong income effects but blacks showing none. A main effects model provided an adequate fit in the 17–44 group with both race and income main effects highly significant.

Among adults 45–64 there also was a significant interaction. Whites showed a strong and consistent income effect. Blacks, however, showed no difference among the three highest income groups but a significantly lower UBDR among the poor. In the oldest age group the interaction was significant at the 0.073 level. Once again whites experienced a strong and consistent income effect but blacks showed a significant difference only between the two highest and two lowest income groups.

Residuals among the estimates for whites were all below 3 per cent. Among blacks, however, some of the residuals were rather large, although no smoothed estimate fell outside the 95 per cent confidence interval based on the corresponding (unsmoothed) cell value.

TABLE A6. Summary of Model-Fitting Procedure: Number of Physician Visits per Year for Those reporting Fair or Poor Health

Model	df	χ^2	p
Saturated model			
ARI = 0*	9	12.32	0.196
Model with ARI = 0†	9	12.32	0.196
AR = 0*	3	18.09	<0.001
AI = 0*	9	33.05	<0.001
RI = 0*	3	5.97	0.113
Main effects model for each age:			
Under 17†	3	1.67	0.642
R = 0*	1	18.81	<0.001
I = 0*	3	37.66	<0.001
17–44†	3	12.21	0.007
45–64†	3	3.27	0.353
R = 0*	1	5.46	0.019
I = 0*	3	5.52	0.138
65 and over†	3	1.14	0.768
R = 0*	1	3.52	0.061
I = 0*	3	16.93	<0.001
Saturated model for age 17–44			
RI = 0*	3	12.21	0.007
Whites			
I = 0*	3	14.66	0.002
P = 1.5P = 2P*	2	0.34	0.844
Blacks			
I = 0*	3	6.20	0.102
Final model†	17	18.13	0.381

* Hypothesis test given specified model.

† These lines give the goodness-of-fit χ^2 for the specified model.

TABLE A7. Summary of Model-Fitting Procedure: Number of Physician Visits per Year for Those Reporting Good or Excellent Health

Model	df	χ^2	p
Saturated model			
ARI = 0*	9	5.03	0.832
Model with ARI = 0†	9	5.03	0.832
AR = 0*	3	39.60	<0.001
AI = 0*	9	64.80	<0.001
RI = 0*	3	1.39	0.707
Main effects model for each age:			
Under 17†	3	3.27	0.352
R = 0*	1	39.97	<0.001
I = 0*	3	125.48	<0.001
P = 1.5P*	1	1.00	0.317
17–44†	3	.66	0.883
R = 0*	1	.44	0.507
I = 0*	3	5.64	0.131
45–64†	3	1.00	0.802
R = 0*	1	2.48	0.115
I = 0*	3	5.65	0.130
65 and over†	3	1.50	0.682
R = 0*	1	5.24	0.022
I = 0*	3	3.32	0.345
Final model†	24	25.56	0.376

* Hypothesis test given specified model.

† These lines give the goodness-of-fit χ^2 for the specified model.

TABLE A8. Summary of Model-Fitting Procedure: Use-Bed-Disability Ratio

Model	df	χ^2	p
Saturated model			
ARI = 0*	9	22.98	0.006
Separate saturated models within each age group:			
Under 17			
RI = 0*	3	8.10	0.044
Whites			
I = 0*	3	90.42	<0.001
Blacks			
I = 0*	3	4.55	0.208
17–44			
RI = 0*	3	3.43	0.330
Model with RI = 0†	3	3.43	0.330
R = 0*	1	20.30	<0.001
I = 0*	3	87.68	<0.001
45–64			
RI = 0*	3	14.84	0.002
Whites			
I = 0*	3	259.53	<0.001
Blacks			
I = 0*	3	14.45	0.002
1.5P = 2P = NP*	2	3.21	0.201
65 and over			
RI = 0*	3	6.96	0.073
Whites			
I = 0*	3	60.53	<0.001
Blacks			
I = 0*	3	20.43	<0.001
P = 1.5P*	1	0.01	0.920
2P = NP*	1	2.31	0.129
Final model†	10	13.51	0.197

* Hypothesis test given specified model.

† These lines give the goodness-of-fit χ^2 for the specified model.

MEDICAL CARE
October 1981, Vol. XIX, No. 1

Improving Hospital Discharge Data: Lessons From the National Hospital Discharge Survey

LINDA K. DEMLO, PH.D.,* AND PAUL M. CAMPBELL, M.S.†

The most recent study by the Institute of Medicine of the reliability of hospital discharge data abstracted from patients' medical records documents the continuing presence of a substantial level of imprecision and error, affirming the findings of earlier studies. The most obvious factors leading to unreliability stem from inadequacies in the face sheets of medical records. Significant improvements could be attained if the discharge summary were routinely used to abstract information on patient disposition and principal diagnosis and if the operative report were routinely used to abstract information on principal procedure. Additional recommendations are offered for developmental activities intended to improve the designation of diagnosis, diagnostic classification schemes and hospital medical records systems.

SEVERAL RECENT STUDIES have questioned the reliability of hospital discharge data and, in particular, information on diagnoses and procedures.[1-6] The latest in a series of studies by the Institute of Medicine examined data collected by the National Hospital Discharge Survey (NHDS), which is conducted by the National Center for Health Statistics (NCHS) and constitutes the only source of nationally aggregated hospital discharge information for the total noninstitutionalized population. The scope of this last study was broadened to permit greater attention to methods for data improvement.

The NHDS is a voluntary survey that has been conducted since 1964 and yields statistics on the utilization of all general and special, nonfederal, short-stay hospitals in the United States and on characteristics of patients who use those services. The information is abstracted primarily from the face sheets of patients' medical records by employees of participating hospitals or by the Bureau of the Census, which acts as an agent for NCHS. The abstracted data are forwarded to NCHS facilities in North Carolina, where they are centrally coded and processed.[7-9]

The division of responsibility for abstracting and coding influences the qual-

---

* Graduate School in Hospital & Health Administration, College of Medicine, University of Iowa, Iowa City, Iowa. At the time of this research, Dr. Demlo was Director, Division of Health Care Services, Institute of Medicine, National Academy of Sciences, Washington, D.C.

† The Orkand Corporation, Silver Spring, Maryland. At the time of this research, Mr. Campbell was affiliated with the Institute of Medicine, National Academy of Sciences, Washington, D.C.

Reprints will not be available from the authors.

Supported primarily by contract no. 282-78-0163, T.O. 4 from the National Center for Health Statistics and also by contract no. 282-75-0437 PM from the Office of Quality Standards and contract no. SSA 600-76-0159 from the Health Care Financing Administration.

0025-7079/81/1000/1030/$01.05 © J. B. Lippincott Co.

ity of the data, because NCHS coders do not have the opportunity to consult the medical record and instead base their code on the handwritten information recorded on the abstract by the person responsible for abstracting in the hospital. The process is further complicated by the inherent complexities of diagnostic information and the wide variation among hospitals in the components, organization and quality of the medical record. This study was designed to take these factors into consideration.

Study Methods

The methods for this study were similar to those of the previous IOM studies, which have been reported elsewhere.[10] Very briefly, the 66 participating hospitals were selected by sampling the NHDS master sample. Within each participating hospital a diagnostic-specific probability sample of NHDS abstracts from calendar year 1977 was drawn, which could be linked to the original medical record within that hospital. The sampling plan was designed to facilitate comparison with some specific diagnoses included in earlier IOM studies and also permitted the aggregation of diagnoses into broader categories so that the reliability of published NCHS statistics could be examined.

The field work was conducted by specially trained Registered Record Administrators (RRAs), who independently reabstracted all patient records selected within participating hospitals, relying primarily on the Uniform Hospital Discharge Data Set definitions. Results of the reabstracting were compared with information on a computerized version of the NHDS abstract, which is part of the NHDS file maintained by NCHS. Discrepancies between the two abstracts were noted. The medical records were then re-examined, along with both the computerized and original hard-copy NHDS abstracts, in an attempt to understand the reasons for discrepancies.

The original study methods were modified slightly to better address the unique features of the NHDS and to build on earlier findings. More emphasis was placed on identifying the location within the medical record of the correct information and, particularly, on assessing the adequacy of the face sheet for abstracting discharge information. More detailed analyses were performed of information on principal procedures. As in previous studies, the diagnostic-specific sampling plan included the total range of the ICDA-8 classification scheme; however, the NHDS study approximated a self-weighting sample more closely than previously.

Nonmedical Data Reliability

As shown in Table 1, the NHDS information was highly reliable for dates of hospital admission and discharge, patient's age or date of birth, sex, race, marital status and patient disposition at the time of hospital discharge. Similar levels of reliability were obtained for most nonmedical data items in previous studies‡ (Table 2). The somewhat lower figures for NHDS data on payment source reflect ambiguity in the procedures for distinguishing between expected principal and additional sources of payment; this has since been alleviated by a routine NCHS computer edit.

Most NHDS discrepancies on dates of hospital admission and discharge, date of birth or age, and sex stemmed from clerical problems, such as transposing code numbers or erroneously recording the year in which the abstracting occurred in place of the year of admission. Most discrepancies

---

‡ The 1977 IOM study of discharge abstracts examined abstracts processed by four private abstract services, reflecting care provided to a national sample of Medicare and Medicaid patients during 1974. The 1977 IOM study of medicare records examined data on utilization by a national sample of Medicare beneficiaries discharged during 1974. The records were generated by the Health Care Finance Administration from claims information submitted by participating hospitals to receive Medicare reimbursement.

TABLE 1. Discrepancy Between NHDS Abstract and IOM Reabstract and the Correct Data Source for Nonmedical Data (Weighted Per Cent)

		Correct Data Source if Discrepant				
Data Item	No Discrepancy	NHDS Abstract	IOM Re-Abstract	Either	Neither	Total
Admission date	99.3	0.4	0.2	0.1	—	100.0
Discharge date	97.9	1.6	0.5	—	—	100.0
Date of birth or age	96.4	1.4	1.7	0.5	—	100.0
Sex	99.5	0.3	0.2	—	—	100.0
Race	92.5	1.1	3.0	3.4	—	100.0
Marital status	95.5	1.7	1.5	1.3	—	100.0
Disposition of patient	92.9	0.4	6.2	0.5	—	100.0

Unweighted N = 3313.

associated with race, marital status and patient disposition were traced to inadequacies of the face sheet of the medical record.

Analyses of Diagnostic Information

Reliability of the NHDS diagnostic information varied considerably, depending on the diagnostic classification and coding schemes that formed the basis of comparison. In Table 3, diagnostic data from all three studies are presented. Four-digit comparisons are quite similar across studies; reliability increases as more inclusive diagnostic groupings are used.

The reliability of individual diagnoses also varied by diagnostic classification and coding. Table 4 shows diagnostic-specific data from all three studies and gives the reliability figures for the most specific, as well as the most inclusive diagnostic groupings within each study. In all studies chronic ischemic heart disease is the most troublesome specific diagnosis. Again, levels of reliability increase as more inclusive diagnostic groupings are used.

TABLE 2. Discrepancies Between the Institute of Medicine Field Team and Original Data Source (Private Abstract, Medicare Record or National Hospital Discharge Survey Abstract) for Nonmedical Data

	Weighted Per Cent with No Discrepancy		
Data Item	NHDS Abstract	Medicare Record*	Private Abstract†
Admission Date	99.3	99.5	99.7
Discharge date	97.9	99.3	99.2
Date of birth or age	96.4	—	97.7
Sex	99.5	99.4	99.1
Payment source	75.6–87.0	—	98.1
Race	92.5	—	—
Marital status	95.5	—	—
Disposition of patient	92.9	—	—

* Reflects claims information for Medicare discharges during 1974.

† Reflects abstract information for Medicare and Medicaid discharges during 1974.

The presence of multiple diagnoses in a single patient influences the reliability of data. Approximately 55 per cent of the cases in the NHDS sample had multiple diagnoses. For those cases, the reliability of the principal diagnosis was 54.7 per cent. This contrasts with a reliability of 73.5 per cent for the cases that did not have additional diagnoses and confirms the findings in the Medicare study. Diagnostic categories whose reliability was notably lowered with multiple diagnoses included diseases of the nervous system and sense organs, diseases of the circulatory system, complications of pregnancy, childbirth and the puerperium and symptoms and ill-defined conditions.

For some cases, particularly when the patient had many medical problems, it was impossible for the field team to determine which diagnosis should be regarded as principal and which data source (the IOM or the original data) should be viewed as correct. For some of these patients, each diagnosis may have contributed equally to the necessity for hospital admission, so that the designation of a single, principal diagnosis would have been artificial and possibly misleading. The per cent of these cases varied, again depending on the level of diagnostic coding refinement. For all records examined in the NHDS study, the figures ranged from 6.5 per cent with four-digit comparisons to 2.0 per cent with the broader NCHS classes. This compares with 4.6 per cent for Medicare data and 10.7 per cent for the private services, both using four-digit coding.

Some uses of hospital discharge data depend on the correct inclusion of a specific diagnosis among the list of diagnoses appearing on the NHDS abstract, regardless of the designation of a principal diagnosis. For almost every diagnosis examined, the likelihood that it would be appropriately included on the NHDS abstract was greater than the likelihood of it being correctly designated as principal. Nevertheless, the more troublesome diagnoses in each analysis tended to be the same—particularly symptoms and ill-defined conditions (Table 5).

TABLE 3. Comparisons Between IOM and Original Data Source for All Diagnoses Combined by Level of Coding Refinement

	Weighted Per Cent with No Discrepancy		
	NHDS Data	Medicare Record*	Private Abstract*
Four-Digit	63.4	58.4	65.2
Three-Digit	73.2	61.7	74.0
CPHA List A	—	—	70.2
DRGs	—	80.8	—
NCHS Categories	86.0	—	—

* See Table 2.

The influence of diagnostic reliability on utilization statistics was estimated by calculating diagnostic-specific hospital discharge rates. Using the broad diagnostic classes appearing in the NCHS Series 13 publications, the NHDS data overestimate discharge rates for slightly more than half the classes in amounts generally ranging from about 3 to 37 discharges per 10,000 population. The largest difference occurs for diseases of the circulatory system, which is caused primarily by the overcoding of chronic ischemic heart disease. Conversely, for slightly under half the diagnostic classes the NHDS data underestimate discharge rates in amounts ranging from 2 to 40 discharges per 10,000 population. The largest undercount occurs for patients with symptoms and ill-defined conditions. Based on IOM findings, the NHDS data for these conditions should have been augmented by 40 discharges per 10,000 population. Next in importance were patients with special conditions and examinations without sickness (or tests with negative findings), for whom the NHDS data should have been increased by about 17 discharges per 10,000 population.

TABLE 4. Comparisons Between the IOM Field Team and Original Data Source by Diagnosis at Differing Levels of Coding Refinement

	Weighted Per Cent with No Discrepancy*					
	NHDS Data		Medicare Record		Private Abstract	
NCHS Diagnostic Category	Four Digit	NCHS Groups	Four Digit	DRGs	Four Digit	CPHA List A
Infective and Parasitic Diseases	60.8	79.4	—	—	—	—
Neoplasms	52.4	74.3	—	—	—	—
Malignant neoplasm of bronchus and lung	—		79.9	79.9	—	—
Carcinoma of the breast	—		—	—	75.7	79.0
Endocrine, Nutritional and Metabolic Diseases	26.3	67.7	—	—	—	—
Diabetes	10.7†	72.1	49.7	56.2	60.8	61.9
Diseases of the Blood and Blood Forming Organs	66.8	72.6	—	—	—	—
Mental disorders	61.7	86.2	—	—	—	—
Neuroses	—		—	—	80.1	80.5
Diseases of the Nervous System and Sense Organs	68.4	79.8	—	—	—	—
Cataract	—		97.4	97.9	94.3	94.2
Diseases of the Circulatory System	53.1	79.7	—	—	—	—
Cerebrovascular accident	43.2	73.6	58.5	84.7	78.4	84.0
Congestive heart failure	68.8	85.4	58.5	60.6	—	—
Chronic ischemic heart disease	37.5	81.4	36.8	38.6	30.2	33.6
Acute myocardial infarction	—	—	67.3	76.1	68.2	93.9
Diseases of the Respiratory System	72.1	92.4	—	—	—	—
Bronchitis	72.2	95.7	89.8	95.7	—	—
Bronchopneumonia	79.6	93.2	75.9	80.5	—	—
Diseases of the Digestive System	70.5	86.2	—	—	—	—
Cholelithiasis/cholecystitis	78.4	94.8	62.8	84.4	73.2	93.8
Intestinal obstruction without mention of hernia	—	—	58.1	68.8	—	—
Diverticulosis of intestine	—	—	86.5	86.9	—	—
Diseases of the Genitourinary System	74.7	87.4	—	—	—	—
Inguinal hernia without mention of obstruction	—	—	96.7	96.7	89.1	92.7
Hyperplasia of the prostate	—	—	87.1	87.1	—	—
End stage renal disease	—	—	—	—	64.9	66.4
Complications of Pregnancy, Childbirth and the Puerperium	64.3	98.0	—	—	—	—
Diseases of Skin and Subcutaneous Tissue	67.9	77.5	—	—	—	—
Diseases of the Musculoskeletal System and Connective Tissue	64.3	86.3	—	—	—	—
Fracture, neck of femur	—	—	70.6	93.4	81.1	97.6
Displacement of intervetebral disc			—	—	77.8	70.9
Congenital Anomalies	49.5	59.8	—	—	—	—
Certain Causes of Perinatal Morbidity and Mortality	74.3	85.3	—	—	—	—
Symptoms and Ill-defined Conditions	54.6	60.5	—	—	—	—
Hypertrophy of tonsils and adenoids	—	—	—	—	97.0	97.0
Accidents, Poisonings and Violence	61.7	89.6	—	—	—	—
Special Conditions and Examinations Without Sickness (or tests with negative findings)	72.8	79.6	—	—	—	—
Newborns	63.0	96.0	—	—	—	—
Deliveries	—	—	—	—	86.5	89.4

* Blanks indicate that the data in the original study were not analyzed in a manner that would permit completing the cell; they do *not* indicate that the relevant diagnoses were omitted from the sample, although in some instances the unweighted n's may have been quite small. All three studies were designed to reflect the total range of diagnoses included in ICDA-8, which was essential to compute net and gross difference rates. The NHDS study placed more emphasis on the broader diagnostic categories than specific diagnosis, whereas the two other studies focused more on the latter.

† This per cent is particularly low because the fourth digit of the NHDS modification is inconsistent with ICDA-8, which was used by the IOM field team.

TABLE 5. Frequency of the IOM Principal Diagnosis Appearing Somewhere on the NHDS Computerized Abstract vs. Frequency of Agreement on Principal Diagnosis by Three-Digit Comparison (Weighted Per Cent)*

IOM Principal Diagnosis	IOM Principal Diagnosis	
	Appears on NHDS Abstract	Agrees with NHDS Principal
Infective and Parasitic Diseases	68.5	62.6
Neoplasms	70.5	57.4
Endocrine, Nutritional and Metabolic Diseases	97.0	71.3
Diabetes	99.8	65.7
All other endocrine, nutritional and metabolic diseases	90.6	84.2
Diseases of the Blood and Blood Forming Organs	92.3	58.2
Mental Disorders	90.3	80.8
Diseases of the Nervous System and Sense Organs	90.0	85.7
Diseases of the Circulatory System	82.5	74.3
Cerebrovascular accident	72.5	61.2
Congestive heart failure	87.7	73.9
Chronic ischemic heart disease	95.1	93.3
All other circulatory diseases	81.3	66.6
Diseases of the Respiratory System	81.6	73.6
Bronchitis	75.3	66.4
Bronchopneumonia	79.4	76.1
All other respiratory diseases	86.4	75.3
Diseases of the Digestive System	85.8	81.0
Cholelithiasis/cholecystitis	90.5	85.8
All other digestive diseases	84.7	79.9
Diseases of the Genitourinary System	87.6	78.1
Complications of Pregnancy, Childbirth and the Puerperium	76.4	75.4
Diseases of the Skin and Subcutaneous Tissue	87.5	73.0
Diseases of the Musculoskeletal System and Connective Tissue	78.4	71.3
Congenital Anomalies	80.0	80.0
Certain Causes of Perinatal Morbidity and Mortality	48.2	20.5
Symptoms and Ill-Defined Conditions	30.4	29.7
Accidents, Posionings and Violence	73.8	67.5
Special Conditions and Examinations without sickness (or tests with negative findings)	50.1	48.6
Newborns	99.2	99.2
Total	80.2	73.1

* In interpreting this table it is important to note that cases were grouped into diagnostic categories according to the diagnosis designated as principal on the IOM reabstract. In previous tables, the categorization is based on NHDS diagnoses. This distinction is important in interpreting diagnostic-specific discrepancy rates for NHDS data, which should be taken from Table 4 and not from Table 5.

Similar comparisons were made for average-length-of-stay (ALOS). For seven of the 19 NCHS diagnostic classes, the NHDS data resulted in ALOS figures that were larger than the IOM by amounts ranging from 0.1 to 1.4 days—the latter was for patients with congenital anomalies. For 9 classes the NHDS ALOS was shorter by amounts ranging from 0.4 days to 15.1 days—the latter for patients with certain causes of perinatal morbidity and mortality. These two extremes were the only classes for which the differences exceeded 1 day.

Most discrepancies in NHDS four-digit diagnostic information arose when there was agreement on principal diagnosis, but disagreement on the appropriate code (referred to as a coding discrepancy). About half the coding discrepancies occurred when the diagnosis on the NHDS abstract was taken directly from the face sheet of the medical record. In these instances the NHDS coders may have correctly coded the face-sheet diagnosis. However, the face-sheet information was incomplete when compared with more refined diagnostic information contained elsewhere in the record. Although disagreement in the selection of the correct principal diagnosis was somewhat less prevalent (referred to as an ordering discrepancy), it again reflected face-sheet deficiencies and usually was caused by designating the first-listed diagnosis on the face sheet as principal. The relative importance of ordering and coding discrepancies was reversed with three-digit comparisons. However, inadequacies of the face sheet remained the most important contributing factor. Inadequacies in the face sheet led to similar problems in the other studies as well.

Analysis of Principal Procedures

The IOM and NHDS abstracts agreed on principal procedure in about 71 per cent of the cases—a slightly lower figure than in previous studies (Table 6). No procedure was coded for more than two thirds of the original abstracts in any study. For those cases, the reliability was higher than for those cases for which a procedure had been coded.

TABLE 6. Comparisons Between the IOM Field Team and Original Data Source on Principal Procedure (Coded to Three Digits)

	Weighted Per Cent With No Discrepancy		
	NHDS Data	Medicare Record	Private Abstract
All cases	71.4	78.9	73.2
Procedure coded	60.1	56.6	66.0
No procedure coded	76.3	89.7	86.7

The reliability of individual NHDS categories of procedures is shown in Table 7. These figures range from a low of 31.9 per cent for obstetric procedures to a high of 98.0 per cent for proctologic surgery.

Most discrepancies on principal procedure arose from disagreement over whether a procedure was important enough to warrant abstracting and coding. This, in turn, was caused by the field team's strict adherence to the UHDDS definition, which caused them to include more procedures than appeared on the NHDS abstract. (The frequency of this type of discrepancy should diminish as the revised UHDDS definitions and guidelines are implemented.[11]) Procedures most likely to be coded by the IOM and omitted by NHDS were those involving contrast radiography, radiotherapies and "other nonsurgical procedures." The other major reason for discrepancy was the failure of the NHDS abstracter to go beyond the face sheet, which frequently did not contain a procedure, to identify a principal procedure found elsewhere in the medical record.

Adequacy of the Face Sheet

In addition to considering the face sheet as a factor contributing to discrepancies, the field team explicitly considered the adequacy of the face sheet for abstracting every information item, regardless of whether a discrepancy occurred. This information is shown in Table 8. The face sheet usually was adequate for nonmedical data, with the exception of patient disposition. The face sheet was much less adequate for medical information. Face sheet information on principal diagnosis was adequate for 47.3 per cent of the records reviewed; the comparable figure for information on principal procedure was only 5.9 per cent.

The location within the medical record of the correct information on patient dis-

TABLE 7. Discrepancy Between the NHDS Abstract and IOM Reabstract and the Correct Data Source for Categories of Procedures Coded to Three Digits by NHDS (Weighted Per Cent)

NHDS Procedures	No Discrepancy	Correct Data Source If Discrepant: NHDS Abstract	IOM Re-Abstract	Either	Neither	Total
Neurosurgery	59.9	—	37.2	2.9	—	100.0
Ophthalmologic surgery	64.0	6.5	16.4	13.1	—	100.0
Otorhinolaryngologic surgery	73.5	2.3	14.1	10.1	—	100.0
Thyroid/parathyroid/thymus or adrenal surgery	75.8	7.2	17.0	—	—	100.0
Vascular/cardiac surgery	56.4	2.2	34.9	5.3	1.2	100.0
Thoracic surgery	69.6	—	30.4	—	—	100.0
Abdominal surgery	82.6	1.8	11.4	4.2	—	100.0
Proctologic surgery	98.0	—	2.0	—	—	100.0
Urologic surgery	70.9	—	21.8	7.3	—	100.0
Breast surgery	62.3	—	24.7	13.0	—	100.0
Gynecologic surgery	58.8	1.1	26.8	13.0	0.3	100.0
Obstetric procedures	31.9	16.8	24.9	26.4	—	100.0
Orthopedic surgery	51.1	1.7	35.2	10.9	1.1	100.0
Plastic surgery	54.8	—	23.2	22.0	—	100.0
Oral/maxillofacial surgery	67.4	14.1	14.6	3.9	—	100.0
Dental surgery	59.3	—	25.2	15.5	—	100.0
Biopsy	54.5	—	35.5	10.0	—	100.0
Diagnostic endoscopy	41.2	—	47.3	11.5	—	100.0
Diagnostic radiography	57.9	13.7	25.9	2.5	—	100.0
Nonsurgical procedures	72.9	14.4	10.6	2.1	—	100.0

position, as well as the medical items, is shown in Table 9. For almost half the cases where the face sheet was inadequate for abstracting patient disposition, the correct information was found on the discharge summary; physician notes were next in importance. The discharge summary contained correct information on principal diagnosis for slightly more than 60 per cent of the comparable cases. The operative report contained the correct information on principal procedure for 52 per cent of the cases where the face sheet had been inadequate; it was followed in importance by

TABLE 8. Adequacy of the Face Sheet for Abstracting NHDS Data (Weighted Per Cent)

Data Item	Adequate	Insufficient	Inaccurate	Total
Admission date	98.8	0.8	0.4	100.0
discharge date	99.2	0.6	0.2	100.0
Date of Birth or Age	96.9	0.8	2.3	100.0
Sex	97.4	1.0	1.6	100.0
Race	91.7	1.0	7.3	100.0
Marital status	95.8	0.2	4.0	100.0
Principal expected source of payment	97.7	0.4	1.9	100.0
Additional expected source of payment	95.9	0.1	4.0	100.0
Patient disposition	24.8	12.0	63.2	100.0
Principal diagnosis	47.3	49.3	3.4	100.0
Principal procedure	5.9	90.6	3.5	100.0

TABLE 9. Location of Correct NHDS Information When Face Sheet Was Inadequate (Weighted Per Cent)

	Data Item		
Portions of Record Used	Patient Disposition	Principal Diagnosis	Principal Procedure
Admit notes	0.1	2.2	—
Consultations	—	0.8	1.1
Discharge summary	48.3	61.4	18.7
Emergency room reports	—	—	0.3
History and physical	0.3	9.8	1.4
Nursing notes	19.3	0.3	1.3
Physician notes	30.9	6.2	7.7
Operative report	—	10.4	52.0
Pathology report	—	1.9	0.3
Other reports	1.1	7.0	17.2
Total per cents	100.0	100.0	100.0

the discharge summary and "other" reports.

Discussion

A review of the findings from all three IOM studies makes obvious the fact that the 1977 data have not significantly improved beyond those examined from 1974. These studies were performed during the time period when four-digit coding of diagnoses was customary. However, the more recent move to a five-digit coding scheme with the introduction of ICD-9-CM, with as yet unestablished levels of reliability, leads to even greater concerns about the reliability of routinely gathered hospital discharge information.

The study committee noted that some improvements in the reliability of hospital discharge data will require fundamental changes in the way that physicians designate diagnoses, the methods by which diagnostic information is classified and the priorities assigned to developing and maintaining good medical record systems within hospitals. These are relatively long-term improvements. However, other improvements can be achieved immediately with relatively simple changes in the processes by which information is abstracted, the design of abstracting forms and their instructions.

Although the current face sheets of medical records are adequate for abstracting most nonmedical information, their inadequacies for abstracting medical data and information on the disposition of the patient at the time of hospital discharge have been clearly established. Significant improvements could be attained if the discharge summary were routinely used to abstract information on patient disposition and principal diagnosis and if the operative report were routinely used to abstract information on principal procedure. Beyond this immediate step, conceptual and methodological improvements are required to further refine all three data items.

With the exception of deaths, information on patient disposition is generally missing from the face sheet and must be inferred from the discharge summary and other portions of the medical record. This information should become a regular component of the face sheet. The concept of patient disposition should be reconsidered and better guidelines developed for its use. If it is intended to describe a patient's continuing health care needs and anticipated service utilization, the current definition will be misleading in several situations. For example, a patient may be discharged to home for a few days because of the nonavailability of a skilled nursing bed, to which he will be admitted as soon as possible. The current discharge categories would result in a discharge to home rather than, more appropriately, to a skilled nursing facility. Similar ambiguity may exist between patients who are discharged home with no continuing care needs and those who will be served by a home care program.

In the absence of significant improvements in the face sheet and discharge summary, it is unlikely that the quality of codes on diagnoses and procedures will attain optimal improvement unless reference is made to the body of the medical record at the time that codes are assigned. Referring to a narrative summary only is inadequate. Accordingly, the committee recommended that NCHS and other data gatherers, who currently code centrally, experiment with requesting coded information directly from hospitals or purchasing coded abstracts from abstract services known to have data of acceptable quality.

If the split between abstracting and coding functions must be maintained, quality-control activities should refer back to the original medical record to determine the extent to which coded information accurately reflects the condition of the patient. Merely checking the internal consistency of the abstracting and coding as separate functions is not adequate.

Reliability of hospital discharge data is associated with the general organization, orderliness and logic of the medical record. The development of a standardized format and guidelines for compiling the face sheet would be one step toward improving the quality and comparability of medical records.

Unreliable hospital discharge data begins with the assignment of a low priority to medical records systems. Hospitals should increase the resources devoted to medical records activities to produce better basic data, particularly in view of the increasingly important decisions being based on such data. Formal communication links should be developed between persons abstracting data and those dealing with the implications of decisions based on hospital discharge data.

Problems of validity or consistency of diagnostic information recorded by the physician suggest that the undergraduate medical curriculum should place greater emphasis on the probabilistic nature of clinical diagnosis and decision making, the extent to which diagnostic styles vary among physicians even apart from issues of diagnostic acumen and accuracy, current classification schemes and the importance of record keeping.

Part of the cause of unreliable diagnostic data is ambiguity in the criteria for designating diagnoses. Improved diagnostic criteria should be developed, preferably through the consensus development activities sponsored by the National Institutes of Health.

Research should be initiated to develop more valid diagnostic classification schemes to acknowledge and appropriately reflect the presence of co-morbid conditions.

Research should begin immediately to assess the reliability of five-digit coding. Any modifications to ICD-9-CM developed for NCHS, Medicare, or any other routine reporting should be reviewed and approved by the U.S. Committee on Vital and Health Statistics to encourage uniform coding and comparability among all hospital discharge abstract systems, both public and private.

Acknowledgment

We wish to acknowledge the significant role of the study committee: Donald C. Riedel, Chairman, Faye Brown, Gerald Charles, James P. Cooney, Paul M. Denson, Sam Shapiro and Joseph Steinberg; and our field team: Mary Rose Dreelin, Maralynn Guenther, Ruth Hayward and John Wiggins. We are also indebted to W. Edward Bacon, Eileen McCarthy, Mary Moien and Earl Bryant of the National Center for Health Statistics and to the participating hospitals and Bureau of the Census Field Staff. The full reports on which this article is based were submitted to the U.S. Department of Health, Education, and Welfare after peer review by the National Academy of Sciences and the Institute of Medicine.

References

1. Institute of Medicine. Reliability of hospital discharge abstracts. Washington, D.C.: National Academy of Sciences, 1977.

2. Institute of Medicine. Reliability of Medicare hospital discharge records. Washington, D.C.: National Academy of Sciences, 1977.

3. New Jersey Department of Health. Reliability of New Jersey hospital discharge abstracts. Trenton: New Jersey Department of Health, Division of Health Planning and Resources Development, 1978.

4. Burford RB JR. The relationship between diagnostic information available at admission and discharge for patients in one PSRO setting: implications for concurrent review. New Haven: Department of Epidemiology and Public Health, Yale University, 1977.

5. California Health Facilities Commission. Reabstracting study, July 1977.

6. JRB Associates. Results of PHDDS reabstracting study. McLean, Va.: JRB Associates, 1980.

7. Simmons WR, Schnack GA. Development of the design of the NCHS hospital discharge survey. PHS Publication no. 1000, series 2, no. 39, September 1970.

8. U.S. Department of Commerce. Hospital discharge survey hospital manual, Washington, D.C.: U.S. Department of Commerce, 1977.

9. Harris KW, Hoffman KL. Quality control in the hospital discharge survey. (DHEW Publication no. (HRA) 76-1342, series 2, no. 68, December 1975.)

10. Demlo LK, Campbell PM, Brown SS. Reliability of information abstracted from patients' medical records. Med Care 1978;16:995.

11. Uniform Hospital discharge data, minimum data set. Report of the National Committee on Vital and Health Statistics. (DHEW Publication no. PHS 80-1157, April 1980.)

MEDICAL CARE
October 1981, Vol. XIX, No. 10

Hospital Medical Staff Organization and Quality of Care:

Results for Myocardial Infarction and Appendectomy

STEPHEN M. SHORTELL, PH.D., AND JAMES P. LOGERFO, M.D., M.P.H.

This article examines the relationships among hospital structural characteristics, individual physician characteristics, medical staff organization characteristics and quality of care for two conditions: acute myocardial infarction and appendicitis. Using data obtained from the Commission on Professional and Hospital Activities (CPHA), approximately 50,000 acute myocardial infarction cases and 8,183 appendectomy cases collected from 96 hospitals in the East North Central Region of the country (Illinois, Indiana, Michigan, Ohio and Wisconsin) were examined. These data were merged with medical staff organization and related data on hospital characteristics obtained from the American Hospital Association. The results indicate that such medical staff organization factors as involvement of the medical staff president with the hospital governing board, overall physician participation in hospital decision making, frequency of medical staff committee meetings and percentage of active staff physicians on contract are positively associated with higher quality-of-care outcomes, independent of the effects of hospital and physician characteristics. Further, the medical staff organization factors appear to be somewhat more strongly associated with higher quality-of-care outcomes than the hospital and physician characteristics. For acute myocardial infarction, higher volume of patients treated per family practitioner and internist and presence of a coronary care unit were also associated with better outcomes. Given the restricted number of conditions studied, the geographically limited sample and the fact that specific variables were not consistently related to quality of care for both conditions, the results are viewed as preliminary. However, they are consistent with and extend other developing findings in this area. They also suggest that more attention needs to be given to the organization of the hospital medical staff and its articulation with the overall hospital decision-making structure and process in attempts to improve outcomes of hospitalization.

THE RELATIONSHIP of hospital medical staff characteristics and organization to costs and quality of care is receiving increasing attention.[1–6] Factors found to be associated with quality of hospital care include hospital structural characteristics such as bed size and teaching status; individual physician characteristics such as specialty and board-certification status; and medical staff organization characteristics such as degree of staff participation in hospital decision making and coordination and control exerted through committees. Unfortunately, most studies have failed to

---

From the Center for Health Services Research, Department of Health Services and Department of Medicine, University of Washington, Seattle, Washington.

Address for reprints: Stephen M. Shortell, Ph.D., Professor and Chairman, Department of Health Services SC-37, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195.

0025-7079/81/1000/1041/$01.25 © J. B. Lippincott Co.

consider all three sets of factors (hospital structure, physician characteristics, medical staff organization) simultaneously in order to examine the relative influence of each.[7] The issue is important because while overall hospital structure and individual physician attributes are relatively difficult to change, medical staff organization processes are much more susceptible to interventions which might improve the quality and efficiency of hospital care.

The purpose of the present study is to examine simultaneously the relationships of hospital, physician and medical staff variables for two conditions: acute myocardial infarction (AMI) and appendicitis. Of particular interest is the examination of whether the medical staff organization variables exert an independent association net of the hospital and physician variables. The existence of such an independent association would suggest that medical staff organization plays an important role as the interface between the hospital and the physician in relation to the quality of care provided. The lack of such an independent association would suggest that the hospital and physician variables are adequately reflecting medical staff organization factors, at least in relation to the quality of care provided for the two conditions under study. If this were the case, it would suggest that little attention needs to be given to medical staff organization itself and that more attention ought to be given to the types of physicians granted staff privileges and the structural characteristics of hospitals granting such privileges.

The above central relationships are examined, taking into account other factors which might be associated with differences in the quality of care provided for the two conditions under study. These include the volume of procedures performed, the degree to which physicians concentrate their work at one hospital, the degree of hospital commitment to quality assurance activities and, in the case of AMIs, the impact of treating patients in a coronary care unit.

The study is primarily exploratory but is guided by a number of propositions suggested in the literature. Given the current state of knowledge, no differential predictions are made for AMI versus appendicitis. Specifically, it is hypothesized that bed size, number of interns and residents per bed and presence of a director of medical education will each be positively associated with higher quality of care. This is based on the assumption that, other things being equal, larger hospitals and those involved in teaching activities have more resources to provide higher quality of care. Board certification as a proxy measure for physician competence is also expected to be associated with higher quality of care. For AMIs it is expected that a higher ratio of internists to family practitioners will be positively associated with higher quality of care. This is based on the assumption that, on the average, internists receive more extensive training in cardiology than family practitioners. Similarly, for appendectomies it is expected that a higher ratio of surgeons to family practitioners will be associated with higher quality of care, assuming that surgeons have more extensive training in surgery than family practitioners. However, in both the AMI and appendectomy cases, physicians' *experience over time with the condition* may offset differences in training.

Based primarily on the findings of Luft et al.,[8] it is also expected that greater volume of procedures performed will be associated with higher quality of care. Given conflicting evidence in the literature,[9–12] no specific prediction is made regarding the effect of treating patients in a coronary care unit, but the available data will enable this relationship to be examined while controlling for a greater number of competing variables and explanations than is presently the case.

The percentage of active staff physicians on contract is hypothesized to be positively associated with higher quality of care, based on the assumption that such a measure at least partially reflects a greater structuring of medical staff organization. Roemer and Friedman[4] have previously found such structuring to be associated with higher quality of care. Concentration of professional activity at one hospital is also felt to be positively associated with higher quality of care on the assumption that such a situation provides more opportunity for peer review, sharing of information among physicians and systematic quality assurance activities. Flood and Scott,[1] for example, have found that the greater the percentage of a surgeon's practice that is concentrated in a given hospital, the better the quality of care.

Physician participation in hospitalwide decision making, measured by such variables as having the medical staff president vote on the governing board and physician participation in the governing board executive committee activities, has also been suggested to be positively associated with higher quality of care.[6,13] The assumption is that such involvement increases the opportunity for awareness and feedback regarding physician practices and provides the opportunity for physicians, administrators and board members to discuss patient care quality issues. Similarly, greater communication via committee meetings among medical staff members is expected to be associated with higher performance. This is because such meetings provide opportunity for learning and peer reinforcement for taking corrective action. Finally, the degree of a hospital's commitment to quality assurance activities is also expected to be positively associated with higher quality of care. Whether or not the hospitals subscribed to an additional medical audit program of the Commission on Professional and Hospital Activities (CPHA) was used as a proxy behavioral indicator of such additional commitment.

Data and Methods

Sample Selection

A 35 per cent proportionately stratified random sample of all hospitals in the East North Central Region (Illinois, Indiana, Michigan, Ohio and Wisconsin) who subscribe to the Commission on Professional and Hospital Activities (CPHA) was selected for study. The East North Central Region was selected because it represents the area of the country within which the highest percentage of hospitals (approximately 45 per cent) subscribe to CPHA. The sample was stratified by bed size, ownership and teaching status, using the following categories: <100 beds, 100–399 beds, and ≥400 beds; short-term voluntary general versus governmentally owned hospitals; teaching hospitals (defined as being affiliated with a medical school or having a residency or internship program) versus nonteaching hospitals. These categories were used in combination to create 12 potential strata. In fact, no hospitals existed in the stratum of less than 100 beds, short-term voluntary general and teaching, nor in the stratum of ≥400 beds, governmental and nonteaching. Of the 131 hospitals sampled, 96 (73.3 per cent) agreed to release their CPHA data for analysis. There was a somewhat higher nonresponse rate for hospitals under 100 beds but little difference by teaching status or ownership.

Data

Given the problems involved in an attempt to measure the overall quality of hospital care, particularly in regard to adjusting for differences in case mix severity, we chose a *condition specific* approach, examining coronary conditions involving myocardial infarction, ischemic heart disease or disorders of heart rhythm (H-ICDA 410.0–416.9) and conditions involving primary appendectomy (H-ICDA 47.0). Of 54,904 coronary care condition patients, 3,745 were excluded due to missing or misrecorded information, leaving 51,159 records available for analysis.

In order to correct for the fact that some patients who were discharged in a few days may have been miscoded as having an

AMI, we excluded those patients discharged alive with approval in five or less days. We note that currently some patients with documented AMI might be discharged as early as the fifth day, but very few such patients would have been discharged this early in 1973, the relevant year of study. These exclusions represented three per cent of all AMI patients.

The remaining group is referred to as the "filtered AMI patients" and represents 97 per cent of the total AMI patients in the study. In order to adjust the expected mortality for each hospital according to case severity on admission, these patients were stratified based on admission systolic blood pressure level and age, two variables which have strong correlation with mortality and which were available from the CPHA data set.[14–16] The six categories based on age and systolic blood pressure readings upon admission to the hospital were as follows: 1) <65 years old and <90 mmHg; 2) ≥65 and <90 mmHg; 3) <65 and 90–99 mmHg; 4) ≥65 and <90–99 mmHg; 5) <65 and ≥100 mmHg; and 6) ≥65 and ≥100 mmHg. The number of patients and the number of deaths within each of the six groups were calculated for each hospital, and a standardized mortality ratio (SMR) was developed, indicating the number of expected versus actual deaths for each hospital based on each hospital's number of patients falling into the respective age and systolic blood pressure categories (see technical appendix for calculation).

For the primary appendectomy study, information on 8,695 patients was originally collected for analysis, but 512 patients were excluded due to missing or misrecorded information, leaving 8,183 patients available for analysis. Since we were interested in the pathologic state of the appendix at operation, we stratified patients based on age and sex, as these characteristics are known to be correlated with the probability of the appendix being normal or perforated.[17] Accordingly, the patients were stratified into the following four groups: 1) males ≤5 years old and ≥65; 2) males, 6–64; 3) females ≤5 years old and ≥65; and 4) females 6–64 years of age. For each hospital, information was available on the total number of patients, number of deaths, number of patients with diagnosis of acute appendicitis with peritonitis (H-ICDA code 540.1) and number of patients with tissue result of no abnormality. We chose to focus on the proportion of normals removed rather than of perforations because removal of a normal is an act that can be avoided without necessarily increasing perforation rates. Furthermore, there are major pitfalls in using case rates of perforation, rather than population-at-risk rates, as an indicator of quality.[18] Thus, a standardized normal tissue removed (SNTR) ratio was computed for each hospital in a fashion similar to that of the standardized mortality ratio for AMI patients. (See appendix.) Thus, the quality-of-care indicators for both the AMI patients and the appendectomy patients have been adjusted for differences in case mix severity and standardized against the experience of other hospitals in the sample.

The hospital, physician and medical staff data are derived from a 1973 national survey of hospital medical staff organization conducted by the American Hospital Association. This survey gathered a variety of data on speciality composition, board-certification status, contractual relationships between physicians and hospitals, governance, decision making, committee structure and related items. For 1973, this information was linked with the CPHA data described previously for the sample of East North Central hospitals to conduct the analyses described below.

Methods

The relative effects of the hospital, physician and medical staff characteristics on the AMI standardized mortality ratio and the standardized per cent normal tissue removed were assessed using multiple-regression analysis. In addition to assessing the effects of the individual variables, indexes of "resource capability," "participation in decision making" and "local staff orientation" were constructed based on previous research and factor analysis of the present data.[6] The variables composing each factor are presented in the appendix. In brief, the resource capability factor represents primarily the ability of the hospital to attract qualified staff and is heavily influenced by hospital bed size and teaching capability. Consistent with previous predictions, it was hypothesized that this factor would be positively associated with better outcomes. Participation in decision making includes measures of physician involvement in hospital wide decision-making bodies and is predicted to be positively associated with better outcomes. Local staff orientation primarily reflects smaller hospitals with fewer physicians per bed and fewer physician members per committee and is hypothesized to be associated with poorer quality of care.

Results are presented for 95 of the 96 hospitals, since data on the SMR were not available for one hospital. In addition, for approximately 12 hospitals, data on the percentage of active staff board certified were estimated using the overall sample mean for per cent of physicians board certified. This substitution yielded results virtually identical to those obtained from deleting these hospitals from analysis. Thus, the substitution results in no bias for the sample overall, and the findings presented below are based on all 95 hospitals.

Results

Acute Myocardial Infarction Results

Table 1 shows the AMI mortality ratios within each of the six age/systolic blood pressure categories. The data indicate significant independent effects of age and systolic blood pressure as well as a strong joint effect. Briefly, controlling for systolic blood pressure on admission, those 65 and over are two to three times more likely to experience death from AMI. Controlling for age, those with systolic blood pressure ≤89 mmHg are approximately three to four times more likely to experience death from AMI than are patients with systolic blood pressure ≥100 mmHg. Looking at the age and systolic blood pressure in combination, patients 65 and over with systolic blood pressure ≤89 mmHg are more than eight times as likely to experience death than those under 65 with systolic blood pressure ≥100 mmHg.

TABLE 1. Mortality Ratios for Filtered AMI Patients Based on Age and Systolic Blood Pressure for 96 Hospitals in East North Central Region

Systolic Blood Pressure	Age	
	< 65 Years	≥ 65 Years
≤ 89 mmHg	0.470	0.828
90–99 mmHg	0.282	0.641
≥ 100 mmHg	0.096	0.334

Table 2 presents the descriptive statistics for all variables examined; the results of the AMI regressions are presented in Table 3. Most of the associations among the variables were in the range of 0.10 to 0.30, with the exception of the correlation between hospital bed size and per cent of active staff physicians board certified (r = 0.41) and the association between bed size and existence of a director of medical education (r = 0.53). As shown, the volume of AMI patients per internist and family practitioner, having the president of the medical staff on the hospital governing board and hospital size are most strongly associated with lower AMI standardized mortality ratios. The unstandardized regression coefficients indicate that a one-unit increase in volume of AMI patients per physician is associated with a 0.014 decrease in the SMR, controlling for or holding constant the other variables in the analysis.

TABLE 2. Descriptive Statistics for Study Variables

Variable	Mean or %	Standard Deviation	N
Standardized AMI mortality ratio	1.06	0.31	95
Standardized per cent normal tissue (appendectomy) removed ratio	1.00	0.68	96
Hospital bed size	242	171.3	96
Number of interns and residents per bed	2 1/100	5.2/100	96
Director of medical education	25.8%	44%	96
Ratio of internists to family practitioners	1.05:1	2.00:1	96
Ratio of surgeons to family practitioners	1.6:1	2.2:1	96
Number of AMI patients per internist and family practitioner*	5.43	4.51	96
Number of appendectomies per surgeon and family practitioner	2.83	2.10	92
Per cent active staff board certified	43.5%	24.4%	84
Per cent active staff on contract	16.8%	11.8%	96
Per cent staff with major professional activity at this hospital	85.4%	24.0%	96
Physicians on executive committee of the hospital governing board	22.1%	41.7%	96
President of Medical Staff on hospital governing board	38.5%	48.9%	96
Number of meetings per medical staff committee per year	11.64	3.52	96
Resource Capability factor†	0.082	3.39	96
Participation in decision making factor†	0.001	2.08	96
Local staff orientation factor†	0.091	4.45	96
Hospital reports having a coronary care unit	63.2%	48.5%	96
Hospital also subscribes to CPHA medical audit program	76.0%	42.9%	96

* This is computed by dividing the total number of AMI patients by the total number of internists and family practitioners on the active staff. Ideally data specific to each internist and family practitioner treating AMI patients are desired but were not available from the existing data. The same is true for number of appendectomies per surgeon and per family practitioner.

† See appendix for further description.

Further analysis also revealed that hospitals seeing 60 or fewer filtered AMI patients per year had a standardized mortality ratio of 1.17, as compared with 1.00 for those seeing more than 60 patients per year ($F = 7.06$, $p \leq 0.009$).

The results also indicate that hospitals having their medical staff president on the governing board experienced standardized mortality ratios 0.108 lower than those of hospitals whose staff president was not on the board, again controlling for other variables in the analysis. Presence of a coronary care unit and per cent of physicians board certified were also associated with a lower AMI standardized mortality ratio, but these relationships were not statistically significant. The per cent of physicians concentrating their activity at one hospital was not found to be significantly associated with lower AMI standardized mortality radios. Whether or not the hospital subscribed to CPHA's additional medical audit program and the ratio of internists to family practitioners were not associated with the AMI standardized mortality ratio. Overall, the above variables explained 17 per cent of the variability in the AMI standardized mortality ratios ($F = 3.10$; $p \leq 0.008$). The amount of variation explained by each variable is largely unaffected by the order in which it entered the equation. The exception is board certification, which when entered last in the equa-

TABLE 3. Individual Variable-Regression Results for AMI Standardized Mortality Ratios

Independent Variable	Unstandardized Coefficient (Standard Error)	Standardized Coefficient*
Volume of acute MI patients per family practitioner and internist	−0.0136† (0.007)	−0.200
President of Medical Staff is on Hospital Governing Board	−0.108‡ (0.063)	−0.171
Hospital bed size	−0.0003‡ (0.0002)	−0.170
Physicians concentrate most of their professional activity at one hospital	0.0017 (0.0014)	0.123
Hospital has a coronary care unit	−0.092 (0.071)	−0.144
Per cent board-certified physicians	−0.0016 (0.0016)	−0.121
$R^2 = 0.174$	$F = 3.10, p \le 0.008$	
$\overline{R}^2 = 0.12$ (Adjusted R^2)		

* The standardized regression coefficients place all variables on equivalent units of measurement based on the standard deviations involved. The magnitude of these coefficients is one way of reflecting the relative importance of the variables in the analysis.

† $p \le 0.05$.

‡ $p \le 0.10$.

tion explained only an additional 0.008 per cent but which in the stepwise regression explained 8 per cent of the variation. The magnitude of the standardized regression coefficients indicates the relative importance of the variables, suggesting that volume, followed by president of staff on governing board and bed size, are the most important predictors.

The impact of the overall factors, shown in Table 4, reveals a moderately strong effect of physician participation in hospital decision making on AMI standardized mortality ratios. The unstandardized regression coefficient indicates that a one-unit increase in this factor is associated with a 0.033 decrease in the AMI standardized mortality ratio. Similarly, a one-unit increase in the volume of AMI patients per internist and family practitioner is associated with a 0.013 decrease in the standardized mortality ratio, and the presence of a coronary care unit is associated with a 0.116 lower AMI standardized mortality ratio. The local staff orientation factor was associated with a higher SMR as predicted, but was not statistically significant. Resource capability was not related to the AMI standardized mortality ratio, nor did it make any difference whether or not the hospital subscribed to CPHA's additional medical audit program. None of the other variables of interest were significantly related. Overall, 18 per cent of the variability in AMI standardized mortality ratio was explained by the above variables ($F = 4.17$; $p \le 0.004$).

When the participation in decision-making factor was forced into the equation first, it explained 7 per cent of the variation in standardized mortality ratios. When it was forced into the equation last, after allowing the other variables to explain as much of the variation as possible, the participation in decision-making factors still explained an additional 5 per cent of the

TABLE 4. Composite Factor Regression Results for AMI Standardized Mortality Ratios

Independent Variable	Unstandardized Coefficient (Standard Error)	Standardized Coefficient
Physician participation in hospital decision making factor	−0.033* (0.015)	−0.225
Volume of acute MI patients per family practitioner and internist	−0.013† (0.007)	−0.187
Hospital has a Coronary Care Unit	−0.116‡ (0.077)	−0.183
Local staff orientation factor	0.010 (0.0085)	0.147
$R^2 = 0.176$	$F = 4.17$, $p \leq 0.004$	
$\overline{R}^2 = 0.134$		

* $p \leq 0.025$.
† $p \leq 0.05$.
‡ $p \leq 0.10$.

variation. An interaction term involving volume times presence or absence of a coronary care unit was also examined but was not significantly related to the standardized mortality ratio.

Appendectomy Results

Table 5 shows the per cent of normal tissue removed within each of the age/sex categories. As indicated, age makes no difference for males, but females between the ages of 6 and 64 are more than twice as likely to have normal tissue removed as females 5 years or younger and 65 or older. The per cent of females between the ages of 6 and 64 with normal tissue removed is approximately twice that of males 6 to 64. The higher rate for females 6 to 64 is primarily due to the number of female conditions, such as ovarian cyst and pelvic inflammatory disease, whose symptoms may be closely related to those of acute appendicitis.

TABLE 5. Per Cent Normal Tissue Removed by Age and Sex for 96 Hospitals in East North Central Region

Age	Male	Female
≤ 5 and ≥ 65	0.088	0.067
6–64	0.083	0.145*

* The increased rate for females is primarily due to the number of female conditions, such as ovarian cyst or pelvic inflammatory disease, whose symptoms may be closely related to those of acute appendicitis.

The results shown in Table 6 are based on 92 hospitals, due to missing data on the variable volume of appendectomies per surgeon and family practitioner. The findings also exclude an "outlier" hospital,* which performed nearly twice as many appendectomies per surgeon and family practitioner per year than other hospitals in the study (21.67 vs. 12.83) and had a standardized per cent normal tissue removed ratio of 3.13, as compared with 2.61 for the next-highest hospital. As shown, the frequency of medical staff committee meetings, the degree to which physicians concentrated most of their activity at the hospital under study, the per cent of physicians on contract and the presence of a director of medical education are each significantly

---

* A 322-bed voluntary community general hospital with no interns and residents.

TABLE 6. Regression Results for Standardized Per Cent Normal Tissue Removed

Independent Variable	Unstandardized Coefficient (Standard Error)	Standardized Coefficient
Frequency of medical staff committee meetings	−0.057* (0.021)	−0.280
Physicians concentrate most of their professional activity at one hospital	−0.0068* (0.003)	−0.240
Percentage of active staff physicians on contract	−0.013† (0.006)	−0.222
Hospital has a director of medical education	−0.246‡ (0.162)	−0.155
Volume of appendectomies per family practitioner and surgeon	0.028 (0.037)	0.088
$R^2 = 0.17$	$F = 3.56; p \leq 0.006$	
$\bar{R}^2 = 0.12$		

* $p \leq 0.005$.
† $p \leq 0.025$.
‡ $p \leq 0.10$.

associated with a lower standardized per cent of normal tissue removed. The unstandardized regression coefficients indicate that a one-unit increase in the frequency of staff committee meetings is associated with a 0.057 decrease in the SMR; a one-unit increase in per cent of physicians concentrating their activity at one hospital is associated with a 0.0068 decrease in the SMR; a one-unit increase in the per cent of active staff physicians on contract is associated with a 0.013 decrease in the SMR; and the presence of a director of medical education is associated with a 0.246 decrease in the SMR. Whether or not the hospital subscribed to CHPA's additional medical audit program, the ratio of surgeons to family practitioners, number of appendectomies performed per surgeon and family practitioner, per cent of physicians board certified and other variables included in the analysis were not associated with the standardized per cent normal tissue removed ratio. The lack of a volume/quality relationship for appendectomies confirms the results of Luft et al.[8] Overall, frequency, of committee meetings, concentration of professional activity at one hospital, per cent of physicians on contract and presence of a director of medical education explained 17 per cent of the variability in the per cent standardized normal tissue removed ratio ($F = 3.56$; $p \leq 0.006$).

It should be noted that the variables were not sensitive to the order in which they entered the equation. For example, whether forced in first or last, volume explained none of the variation in per cent standardized normal tissue removed. In contrast, frequency of committee meetings explained 7 per cent of the variance even when forced in last and also explained 7 per cent when forced into the equation first.

Analysis of the overall factors involving resource capability, medical staff participation in decision making and local staff orientation revealed no significant associations. The only variable that continued to be significantly associated with a lower standardized per cent normal tissue removed ratio was concentration of major professional activity at the hospital under study. The higher the percentage of physicians indicating that they concentrated their activity at one hospital, the lower the per cent standardized normal tissue removed ratio.

Discussion

Differences in mortality in coronary care units are well documented.[19–21] Less well documented are the factors that might be associated with these differences. Bloom and Peterson found that coronary care units with full-time medical directors experienced better performance.[19] Stross et al.[21] found that units seeing fewer than 60 AMIs a year experienced *twice* the mortality of those seeing 60 or more per year. Others have identified the critical elements in treating CCU patients, including the positive impact of educating CCU nurses in problem recognition and response.[22,23] Policy changes permitting nurses to intervene immediately have also been associated with declines in mortality.[24] On a different level, the very effectiveness of CCUs in treating AMI patients has been questioned relative to at-home treatment and treatment in regular hospital beds.[9–12]

The present results suggest that the presence of a CCU does make a difference in the standardized mortality ratios of AMI patients. Hospitals with a CCU had lower standardized mortality ratios than those without a CCU, even after controlling for hospital size, volume and related variables. However, for those hospitals with a CCU, it was not possible to analyze whether those patients in the CCU did better than those treated in private or semiprivate rooms.

Further, the results suggest that volume of AMI patients treated is, indeed, a significant factor, with those treating 60 or fewer AMI patients a year experiencing a 17 per cent greater standardized mortality ratio than those treating more than 60 a year. More relevant is the fact that number of AMI patients treated *per* internist and family practitioner is significantly related to lower standardized mortality ratios, even after taking other variables into account. This finding suggests that recent research linking volume of surgical operations to postoperative mortality may also be part of a more generalized phenomenon in which the amount of clinical experience with a given condition within a given year plays an important role in nonsurgical cases as well.[8] If supported by other studies, the results suggest some important implications for grouping or clustering CCU services around those hospitals with sufficient volume to maintain high performance outcomes. Under the guidance of health systems agencies and faced with other external pressures, hospitals are already beginning to group or cluster their obstetric and pediatric services. However, treatment of AMI patients raises a somewhat more difficult set of issues in terms of travel time to receive treatment factors. Thus, a policy of reducing the number of individual hospitals with CCUs would also depend in part on the availability of competent emergency medical transport services available, particularly in rural and semirural areas. The question also arises as to whether volume or the existence of a CCU is the more important factor. Results from Table 4 suggest that both are about equally important in terms of their association with standardized mortality ratios. However, analysis of covariance involving only volume, presence or absence of a CCU and standardized mortality ratios revealed a stronger effect of volume ($F = 7.91$; $p \leq 0.006$) than of presence or absence of a CCU. ($F = 2.70$; $p \leq 0.104$.)

The present results involving AMI standardized mortality ratios also suggest the important role played by medical staff organization, particularly in regard to participation in hospital decision making. Whether or not the president of the medical staff was on the hospital governing board and an overall measure of participation, which included also whether or not other medical staff officers were on the governing board, and whether there were physicians on the governing board executive committee, were significantly related to lower standardized mortality ratios, taking into account other possible explanatory factors. In fact, the present results suggest that the overall medical staff participation in decision making factor is the single most important variable associated with lower standardized mortality ratios, more significant than either the volume of AMI patients per internist and family practitioner or the presence or absence of a coronary care unit. Other variables commonly believed to be associated with better performance, such as hospital bed size, degree of teaching activity and per cent board-certified physicians, were not as significantly associated with lower standardized mortality ratios when the above variables were taken into account. Each of them, however, was in the expected direction.

The literature suggests wide variation in the removal of histologically normal appendices.[25–28] Yet, again, there is little documentation of the factors that might be associated with the amount of normal tissue removed. Neutra's research suggests that with a more careful history and exam and close observation for several hours, hospitals can reduce the rate of normal tissue removed without any increase in the rate of perforation.[25] The present analysis suggests some medical staff organization factors which may be associated with and may facilitate the kinds of behavior suggested by Neutra. Specifically, frequency of medical staff committee meetings, concentration of physician activity around one hospital, per cent of active staff physicians on contract, and presence of a director of medical education are each associated with a lower standardized per cent normal tissue removed ratio. Frequent medical staff committee meetings promote communication and exchange of information among staff members, increasing the probability of implementing changes in treatment practices, which may improve the process and/or outcomes of care. Lending support to this notion, Morlock et al. recently found a significant association between frequency of medical staff committee meetings and lower overall hospital death rates after adjusting for patient diagnostic mix.[29]

Concentration of professional activity at one hospital offers greater opportunity for peer review and establishment of new practice habits and patterns. For example, in early quality-of-care research, Morehead found that those physicians who spent most of their time (greater than 75 per cent) treating Health Insurance Plan (HIP) of Greater New York patients provided a higher quality of care than those scattered throughout several locations or those seeing HIP patients only part-time.[30] Flood and Scott have found that the greater the percentage of a surgeon's practice that is concentrated in a given hospital, the better the quality of patient care.[1]

Having a greater percentage of physicians on contract also affords an opportunity for closer physician involvement in hospital treatment practices and facilitates the opportunity to initiate and implement improvements in care. The present results in this regard are also supported by previous work which found positive associations between measures of physicians on contract and a variety of quality-of-care measures.[4] Finally, the presence of a director of medical education may provide the kind of systematic leadership and direction which, when combined with the greater emphasis on teaching implied by such a position, may stimulate better treatment practices.

The above findings and interpretations are only suggestive. Variables associated with differences in outcomes for AMI were not the same as those associated with differences in outcomes for appendicitis, although both sets of variables contained measures of medical staff organization concepts. Findings for these conditions should be replicated. Subsequent research should also examine other conditions as well as hospitalwide measures of quality in order to extend present results. Further specification of volume–quality relationships with regard to concentration of admissions and operations around smaller numbers of physicians would be a useful avenue of inquiry. Particularly useful would be studies of the impact of planned changes in medical staff organization and hospital–physician relationships on quality of care that would include data from hospitals not introducing such changes. Variations in implementation of quality assurance programs might be a source for such studies. The impact of regulation (rate review, certificate of need, and so on) on physician–hospital relationships and medical staff organization in relation to both cost and quality considerations is also a fertile area for study. Finally, it is of interest to note that the relatively aggregate measures used in the present research explained approximately 15 per cent of the variation in quality of care. Thus, examination of medical staff and hospital organization at a less aggregate or more micro level (such as the actual organization and staffing of coronary care units) would be likely to add to the explanation of differences in quality of care across hospitals.

While inferences drawn from the above results can, in a strict sense, only be generalized to other CPHA-subscribing hospitals in the East North Central Region of the country, the findings are suggestive of some important medical staff influences which may hold implications for many U.S. hospitals. Of particular interest is the fact that, for the conditions and hospitals studied, neither hospital structural characteristics such as bed size and degree of teaching involvement nor individual physician characteristics such as specialty composition were strongly related to better performance. Rather, medical staff organization characteristics, such as degree of physician participation in hospital decisionmaking, frequency of committee meetings, concentration of activity in one hospital and percentage of physicians on contract, were found to be more strongly associated with superior performance, although not consistently so for both conditions. In addition, for AMI patients, the presence of a coronary care unit and the volume of AMI patients per internist and family practitioner were related to superior performance. It appears that simply having "good" people (e.g., well-trained, board-certified physicians) and having these people affiliate with "good" hospitals (e.g., large-size, teaching-oriented hospitals) is not sufficient to guarantee superior results. Rather, the actual organization of the medical staff itself, as reflected partly by participation in hospitalwide decision-making bodies and communication through committee meetings, appears to be most strongly associated with the quality of care provided. Thus, activities aimed at changing medical staff organization and articulating its relationship and involvement with the overall hospital organization and decision-making process may be a key factor in improving the quality of hospital care.

Acknowledgment

This research was supported by the National Center for Health Services Research under Contract Grant # HS 01978 to the Center for Health Services, University of Washington, and also supported by the Office of Research Affairs, American Hospital Association and by Group Health Cooperative of Puget Sound. Appreciation is expressed to Walt Wood, Commission on Professional and Hospital Activities (CPHA), Ann Arbor, Michigan, for his assistance in expediting provision of the data for analysis. Appreciation is also expressed to Ray Brasto and Paul Haley for the computer programming work associated with these analyses. The analysis, interpretations, and conclusions are the sole responsibility of the authors.

References

1. Flood A, Scott WR. Professional power and professional effectiveness: the power of the surgical staff and the quality of surgical care in hospitals. J Health Soc Behav 1978;19:240.

2. Pauly MV. Medical staff characteristics and hospital costs. J Hum Resour 1978;13(Suppl):77.

3. Rhee SO. Factors determining the quality of physician performance and patient care. Med Care 1976;14:733.

4. Roemer M, Friedman J. Doctors and hospitals: medical staff organization and hospital performance. Baltimore: Johns Hopkins Press, 1971.

5. Shortell SM, Becker S, Neuhauser D. The effects of management practices on hospital efficiency and quality of care. In: Shortell SM, Brown M, eds. Organizational research in hospitals. Chicago: Blue Cross Association, 1976:90–107.

6. Shortell SM, Getzen T. Measuring hospital medical staff organizational structure. Health Serv Res 1979;14:97.

7. Palmer RH, Rielly MC Individual and institutional variables which may serve as indicators of quality of medical care. Med Care 1979;17:693.

8. Luft H, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med 1979;388:1364.

9. Mather HG, Pearson NG, Reed KLQ, et al. Acute myocardial infarction. Home and hospital treatment. Br Med J 1971;3:334.

10. Dellipiani AW, Kolling WA, Donaldson RJ, et al. Teesside Coronary Survey: fatality and comparative severity of patients treated at home, in the hospital ward, and in the coronary care unit after myocardial infarction. Br Heart J 1977;39:1172.

11. Hill JC, Hampton JR, Mitchell JRA. A randomized trial of home versus hospital management for patients with suspected myocardial infarction. Lancet 1978;i:837.

12. Astvad K, Fabricius-Bjerre N, Kjaerulff J, et al. Mortality from acute myocardial infarction before and after establishment of a coronary care unit. Br Med J 1974;1:567.

13. Shortell, SM, Evashwick C. The structural figuration of U.S. hospital medical staffs. Med Care 1981;19:419.

14. Helmers C. Short and long term prognostic indices in acute myocardial infarction. ACTA Med Stand (Suppl SSS), 1973.

15. Schor S, Shani M, Madar B. Factors affecting immediate mortality of patients with acute myocardial infarction: a nationwide study. Chest 1975;68:217.

16. Peterson DR, Thompson DV, Chinn N. Ischemic heart disease prognosis: a community-wide assessment. JAMA 1972;219:1423.

17. Gilmore OJA, Browett JP, Griffin PH, et al. Appendicitis and mimicky conditions. Lancet 1975;ii:421.

18. LoGerfo JP, Brook, RH. Evaluation of services and care, Chapter 13. In: Williams SJ, Torrens PR. Introduction to health services. New York: John Wiley and Sons, 1980:361–363.

19. Bloom BS, Peterson OL. End result, cost and productivity of coronary care units. N Engl J Med 1973;288:72.

20. Klaus AP, Sarachek MS, Greenberg D, et al. Evaluating coronary care units. Am Heart J 1970;79:471.

21. Stross JK, Willis PW, Reynold EW Jr., et al. Effectiveness of coronary care units in small community hospitals. Ann Intern Med 1976;85:709.

22. Clipson CW, Wehrer JJ. Planning for cardiac care: a guide to the planning and design of cardiac care facilities. Ann Arbor: The Health Administration Press, 1973.

23. Stross JK, Bellfy LC. Health Lung 1979;8:318.

24. Killip T, Kamball JT. Myocardial infarction in a coronary care unit: a two year experience with 250 patients. Am J Cardiol 1976;20:474.

25. Neutra RR. Appendicitis: decreasing normal removals without increasing perforations. Med Care 1978;16:956.

26. Lichtner S, Pflanz M. Appendectomy in the Federal Republic of Germany: epidemiology in medical care patterns. Med Care 1971;9:311.

27. Thomas EJ, Mueller CB. Appendectomy: Diagnostic criteria and hospital performance. Hosp Pract 1969;4(4):72.

28. Wennberg J, Gittelsohn A. Small area variations in health care delivery. Science 1973;182:1102.

29. Morlock L, Nathanson CA, Horne SD. Organizational factors associated with quality of care in 17 general acute care hospitals. Paper presented at the annual meeting of the Association of University Programs in Health Administration, Toronto, Canada, May 6, 1979.

30. Morehead MA. Quality of medical care provided by family physicians as related to their education, training and methods practice. New York: HIP of New York, 1958.

Appendix

A. Calculation of Standard Mortality Ratio (SMR) for AMI Patients and Standardized Normal Tissue Removed Ratio (SNTR) for Appendectomy Patients.

1. Standardized Mortality Ratio (SMR) =

$$\frac{\sum_{i=1}^{6} d_i}{\sum_{i=1}^{6} n_i \, (P_i)}$$

where:

d_i = each hospital's number of AMI deaths in each age, systolic blood pressure category

n_i = each hospital's number of filtered AMI patients in each age, systolic blood pressure category

P_i = total number of deaths in each age, systolic blood pressure category for *all* study hospitals *divided by* the total number of patients in each age, systolic blood pressure category for *all* study hospitals.

2. Standardized Normal Tissue Removed Ratio (SNTR) =

$$\frac{\sum_{i=1}^{4} t_i}{\sum_{i=1}^{4} n_i \, (P_i)}$$

where:

t_i = each hospital's number of patients with normal tissue removed in each age, sex category

n_i = each hospital's number of appendectomy patients in each age, sex category

P_i = total number of patients with normal tissue removed in each age, sex category for *all* study hospitals *divided by* total number of appendectomy patients in each age, sex category for *all* study hospitals.

B. *Factor Analysis Results*

1. *Resource Capability Factor*

Variables	Factor Loadings	Communality
Hospital bed size	0.82	0.82
Existence of a director of medical education	0.67	0.55
Number of services and benefits provided to medical staff	0.51	0.51
Per cent of active staff board certified	0.45	0.63
Number of interns and residents/bed	0.36	0.77
Per cent of active staff who are radiologists and pathologists	-0.41	0.81

Eigenvalue = 4.96.
Per cent of variance explained = 30.3 per cent.

2. *Participation in Decision Making Factor*

Variables	Factor Loadings	Communality
Other medical staff officers are on governing board	0.84	0.81
Physicians are on executive committee of governing board	0.44	0.37
President of medical staff is on hospital governing board	0.29	0.41

Eigenvalue = 1.39.
Per cent of variance explained = 8.5 per cent.

3. *Local Staff Orientation Factor*

Variables	Factor Loadings	Communality
Per cent of active staff who are radiologists and pathologists	0.82	0.81
Per cent of active staff who are generalists (i.e., family practitioners and pediatricians)	0.32	0.29
Number of active staff physicians per bed	-0.31	0.90
Per cent of active staff board certified	-0.36	0.63
Number of medical staff members per committee	-0.36	0.64
President of medical staff is elected by the staff	-0.43	0.30

Eigenvalue = 1.60.
Per cent of variance explained = 9.8 per cent.

MEDICAL CARE
October 1981, Vol. XIX, No. 10

Communications

Noninvestigational Use of Unapproved Drugs—Experience From the Scandinavian Countries

ELINA HEMMINKI, M.D.

In the Scandinavian countries it is possible to use unapproved drugs for treatment purposes with special permission from the drug control authorities (individual licenses). The Scandinavian practice and the problems connected with it are described. The basic principle underlying the laws and regulations has been and is the same in each country: unapproved drugs should be used only in exceptional cases for special reasons. However, the practice has become different: many individual licenses have been granted, giving physicians much greater opportunity to prescribe drugs of their choice than the apparently tight Scandinavian licensing laws and practices would suggest. The great number of individual licenses has become an administrative problem and has also undermined the principle underlying the law.

THE PRESENT DRUG LAWS in the United States allow interstate sales only of those drugs which have been approved (licensed) by the Food and Drug Administration (FDA). However, according to the report of the U.S. Senate Committee on Labor and Human Resources,[1] FDA, under the guise of approving drugs for investigational purposes, has permitted small numbers of patients with serious diseases to use unapproved drugs. The 1979 Drug Regulatory Reform Act[1] proposes that with certain criteria, use of unapproved drugs should be allowed in the treatment of small numbers of patients.

In the Scandinavian countries it has traditionally been possible to use unapproved drugs for treatment purposes with special permission from the drug control authorities. (These permissions will be called "individual licenses" henceforth.) The purpose of this article is to describe the Scandinavian practice and the problems connected with it.

The data on which the article is based consist of laws, statutes and other official documents; archive material; interviews and previous studies. (A list of laws and other official documents is available from the author.)

The Laws and Practice

Laws requiring drug licensing came into force in 1930 in Norway, in 1934 in Sweden, in 1954 in Denmark and in 1964 in Finland. The general features of the present licensing laws and practices have been

From the Finnish Medical Research Council, The Academy of Finland, and Department of Public Health Sciences, University of Tampere, Finland.

Supported by The Academy of Finland and the Finnish Foundation for Alcohol Studies.

Address for reprints: Elina Hemminki, Projektet Psykofarmaka i Norden, The Finnish Foundation for Alcohol Studies, Kalevankatu 12, SF-00100 Helsinki 10, Finland.

0025-7079/81/1000/1056/$00.75 © J. B. Lippincott Co.

described earlier.[2,3] In Denmark and Sweden the option for individual licenses was stated in the laws from the beginning of drug licensing. In Norway this option was introduced in 1941, and in Finland in 1969. But at least in Finland, individual licenses were permitted before that date.

The basic contents of the laws and regulations have been and are the same in each country: unapproved drugs should be given only for special reasons, and to obtain such drugs a special application must be sent to drug control authorities. In Finland, Norway and Sweden the special reasons must be health or health care related. In Norway and in most years in Sweden (1934–1952, 1963–present), it has been further stated that the reasons should be given for each patient. In the other years in Sweden, individual licenses were permitted for individual drugs for specified physicians, who could use them for several patients. In Finland and Denmark the laws have been more vaguely formulated, and whether individual licenses should be given per patient, physician or institution has not been clearly specified.

The laws and other official documents give an impression of strict rules, but the practice has been different. In each country applications for individual licenses have been accepted more or less automatically, and only a few have been rejected. One or two people have handled the applications for individual licenses, usually in addition to their other duties. Comparison with the large number of applications shows that consideration of each application cannot have been thorough. In Norway until 1977, there was not much difference for physicians between writing an application for an individual license and writing a normal prescription. Only since 1977 has a special statement justifying the

TABLE 1. Number of Individual Licenses in the Scandinavian Countries*

Year	Denmark	Finland	Norway†	Sweden‡
1949	NA	NA	?	2,100
51	NA	NA	?	3,242
55	?	NA	?	3,086
57	?	NA	2,120	7,539
59	?	NA	2,433	9,798
61	?	NA	3,280	20,486
63	?	NA	3,200	23,525
65	?	?	3,272	11,013
67	?	?	3,984	3,133
69	?	?	4,284	3,417
71	?	?	4,738	6,141
73	?	?	6,546	14,922
74	960	917	7,757	19,502
75	973	872	~ 9,800	20,803
76	1,668	1,244	~12,400	23,350
77	1,971	1,424	16,189	27,802
78	?	1,959	[illegible]	33,409
79	?	?	12,929	39,594
Population in 1978, in 1000s	5,104	4,752	4,059	8,278

* The Danish and Finnish figures concern number of applications for individual licenses, the Norwegian and Swedish figures licenses granted. NA = not applicable; ? = not available.

† Includes individual licenses for drugs used in clinical trials.

‡ Includes individual licenses for special foods, see text.

Sources: 4–12.

need of the unapproved drug been required. Pharmacists have taken care of the application for the individual license: they have ordered the drug, and a copy of the prescription has been sent to the drug control authorities, who have decided about an individual license. In the other countries the practice has been similar, but statements about need were already required as long ago as 1956 in Sweden.

The Danish, Finnish and Norwegian laws have stated that only small amounts of drugs should be given through each individual license. In Finland this was explicitly stated by the drug control authorities: for outpatients it was limited to the amount necessary for 3 months of treatment between 1969 and 1975, and to a year since 1975, while amounts for hospital use could be larger. Nowadays, most of the individual licenses in Finland are given to hospitals. The allowable amounts have not been defined as exactly in the other countries. Yet, apparently, at least those individual licenses given to hospitals and not specified per patient have permitted large amounts of drugs. Such unspecified individual licenses for hospitals have also been given in Norway and Sweden, even though they are against the spirit of the regulations about individual licenses. In Norway such licenses have accounted for about a quarter of all individual licenses in the 1970s.[4] It is true that in Norway the drug control authorities have been given a general title to exempt the licensing regulations. Apparently, this title has been applied when these unspecified individual licenses have been granted.

The numbers of individual licenses show that they have not been rare exceptions (Table 1). The figures are those given by the control authorities themselves. They are not comparable between the countries, because the types of products covered by individual licenses and the amounts allowed on one license vary from one country to another. For example, many individual licenses for special foods are included in the Swedish data, and herbals and homeopathics are included in the Norwegian data. In Sweden there were 19 individual licenses for special foods (0.6 per cent of the total) in 1967, 2,816 (19 per cent) in 1973, and 12,848 (46 per cent) in 1977.[13] On the basis of the data given by Christiansen,[4] I estimated that in 1977 licenses for herbals and homeopathics accounted for a fifth of all individual licenses in Norway. In the statistics one individual license may refer to a small amount of a drug for a short treatment of one patient or to a large amount for the treatment of several patients in a hospital. (The individual licenses given to specified physicians in the period of 1952–1963 in Sweden were apparently recorded per patient, and this explains the high number of individual licenses during that period.) Within each country the recording practices do not vary much in a short time period, and it may be concluded from Table 1 that in each country the number of individual licenses has increased during the 1970s. In 1979 the control authorities in Norway had to handle as a mean 60 requests per working day, and in Sweden some 180 requests.

Reasons for Individual Licenses

I have interpreted the spirit of the Scandinavian drug laws and determined that the following have been "acceptable reasons" for individual licenses: 1) treatment of a disease which occurs only rarely in the country; 2) continuation of treatment of a person only visiting the country; 3) use of a drug regarded as undesirable for general use but considered beneficial for some patients. As "unacceptable reasons" the following three can be listed: 4) an application for the drug to be licensed has been made, and individual licenses are applied for to pressure the control authorities and to promote the drug; 5) the control authorities have rejected or would reject an application for the drug, so individual licenses are used to avoid conflict (a means of getting around the licensing regulations); 6) the producer of the drug does not want to market the drug for commercial or technical reasons.

The numbers of individual licenses and especially their increase suggest that the acceptable reasons cannot explain all of them. Part of the increase in the individual licenses in Sweden may be explained by the licenses for special foods (see above). But even when these are excluded, the number of individual licenses has been high and is increasing in Sweden. A study by Christiansen[4] allowed me to estimate the reasons for individual licenses in Norway in the 1970s. She took a one-third random sample of individual licenses in 1974 and 1977. In her report the drugs for which individual licenses were most commonly applied (66 per cent of all applications in 1974 and 79 per cent in 1977) are listed and discussed. I reanalyzed her raw data and grouped the individual licenses by the status of application for a general license (Table 2). Table 2 can be interpreted and simplified in the following way: about one third of the individual licenses may have represented pressure on control authorities and premarketing promotion (group 1). Another third may have been for the purpose of getting around licensing regulations (group 2). (The individual licenses for herbals and homeopathics are included in this group, because I believe that the Norwegian control authorities have considered them ineffective.) These are crude estimates, and a small proportion of the individual licenses in groups 1 and 2 have been for acceptable reasons. About 15 per cent of the individual licenses were for drugs which the producers did not want to market (group 3). The actual numbers of individual licenses increased from 1974 to 1977 in each group, because the total number more than doubled during this period.

Discussion

The large number of individual licenses, the limited resources reserved for their handling within control authorities, the individual licenses to hospitals which were not per patient, scattered data from several sources and my analysis from the Norwegian data (Table 2) show that the individual licenses in the Scandinavian countries have not reflected their original purposes. They have given physicians a much greater

TABLE 2. Individual Licenses by the Status of the Application for General License; Most Commonly Applied Drugs in Norway in 1974 and 1977*

	%			
	1974		1977	
1. Application under consideration	32		35	
2. Application rejected	12	34	12	37
Drug withdrawn from market by control authorities' decision	1		2	
No application; drug likely to be ineffective†	21		23	
3. Drug withdrawn from market by decision of the firm	2	14	1	16
No application, drug likely to be effective	12		15	
4. No application, efficacy unknown	13	20	9	12
Cannot be classified	7		3	
Total	100		100	
Number of individual licenses	(1605)		(4175)	

* Based on the data provided by Christiansen (1979), see text. Drugs prescribed by dentists and veterinarians are excluded.

† Includes herbals and homeopathics.

opportunity to prescribe drugs they choose than the apparently tight Scandinavian licensing laws and practices would suggest. To the drug industry it has offered a way of putting pressure on the control authorities while the application for licensing is under consideration and a way of promoting the drug before approval. I do not know whether firms have received direct financial profits from individual licenses. In Norway individual licenses accounted for 2.4 per cent of the value of the total sales of drugs in 1977.[5,14]

In recent years the administrative burden caused by the increasingly large number of applications for individual licenses has drawn attention to them. To alleviate the situation a proposal has been made in Sweden to allow certain institutions to use unapproved drugs officially without applying for permission for each patient (a practice widely used in the other Scandinavian countries).[13] In Norway new administrative measures and information aimed at reducing the number of individual licenses have been put forward, and since 1977 the number of individual licenses has started to decrease.

Yet the great number of individual licenses is of concern also as a matter of principle. The purpose of drug licensing in the Scandinavian countries is to ensure that only effective, safe and reasonably priced drugs are on the market. Certain bodies within the drug control authorities have been charged with making these judgments. The system of individual licenses was created to allow exceptional use of unapproved drugs for special (health-related) reasons. If these principles are considered poor they should be officially changed, and not breached in practice so that part of the drug consumption is determined by a different principle, i.e., the judgment of practicing physicians.

My conclusion is that the principle of the individual licenses could be retained, but that present practice should be changed to correspond to the spirit of the law, with a reduction in the numbers of individual licenses. A special problem is posed by individual licenses for useful drugs, which drug firms for one reason or another have not tried to bring onto the market. Soliciting of applications for such drugs, or for unpatented drugs, their production by noncommercial institutions, might help.

Acknowledgments

This study is a part of the project "Psychotropics in the Nordic Countries 1950–1977." In Finland the study was supported financially by the Academy of Finland and the Finnish Foundation for Alcohol Studies. I wish to thank Kettil Bruun, Eivind Falkum and John Lilja for their help in preparing this article.

References

1. Committee on Labor and Human Resources, United States Senate. The Drug Regulation Reform Act of 1979. Washington: U.S. Government Printing Office, 1979.

2. Wardell WM, ed. Controlling the use of therapeutic drugs. Washington: American Enterprise Institute for Public Policy Research, 1978.

3. Hemminki E, Falkum E. Psychotropic drug registration in the Scandinavian countries: the role of clinical trials. Soc Sci Med 1980;14A:547.

4. Christiansen A. Forskrivning av ikke-registrerte preparater. University of Oslo, 1979.

5. Nordiska läkemedelsnämnden. Nordic Statistics on Medicines 1975–1977. Helsinki, 1979.

6. National Board of Health (Finland). Unpublished data, 1979.

7. Falkum E, Jensen TØ, Bomann-Larsen P. Kontroll med legemidler i Norge. University of Bergen, 1979.

8. Lilja J. Läkemedelskontrollen i Sverige 1950–1978. University of Uppsala: unpublished manuscript, 1979.

9. The National Board of Health and Welfare (Sweden). Archives of the Board of Drugs, 1955–1970.

10. Apoteksbolaget. Statistiska sammanställningar 1979. Stockholm, 1980:37.

11. Central Statistical Office of Finland. Statistical yearbook of Finland 1979. Helsinki, 1980.

12. Statens legemiddelkontroll. Nytt fra Statens legemiddelkontroll, Nr 2, 1980:2.

13. Socialdepartementet. Den statliga läkemedelskontrollen. DsS 1978:12. Stockholm, 1978.

14. Statens legemiddelkontroll. Nytt fra Statens legemiddelkontroll, Nr 2, 1978:1.

C 5170-1

Medical Care

November 1981
Vol. 19 No. 11
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

DRG-Based Reimbursement: The Use of Concurrent and Retrospective Clinical Data

A Case-Mix Method for Developing Health Planning Criteria for Hospital Services

Why Do Nursing Home Costs Vary? The Determinants of Nursing Home Costs

Statistical Analysis of an Attitude Scale to Measure Patient Satisfaction With Medical Care

Life Satisfaction and Chronic Disease: Exploration of a Theoretical Model

Refinements in the Measurement of Health-Specific Locus-of-Control Beliefs

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
The Rand Corporation

Gordon H. De Friese, Ph.D.
University of North Carolina

Brant E. Fries, Ph.D.
Yale University

Paul M. Gertman, M.D.
Boston University Medical Center

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Joel W. Kovner, Dr.P.H.
Kaiser Permanente Medical Care
Programs, Los Angeles

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

Steven A. Schroeder, M.D.
University of California,
San Francisco

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $40.00; special price U.S. students, $23.00; all other countries except Japan, one year, $50.00. Single copies $5.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1981 by J. B. Lippincott Company. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

November 1981
Volume 19, Number 11

CONTENTS

POLICY FORUM

What Do Patient Statements Regarding Doctors and Medical Care Signify? *Naomi Breslau* 1069

ORIGINAL ARTICLES

DRG-Based Reimbursement: The Use of Concurrent and Retrospective Clinical Data. *Cynthia Barnard and Truman Esmond* 1071

A Case-Mix Method for Developing Health Planning Criteria for Hospital Services. *James Greenberg and Roger Kropf* 1083

Why Do Nursing Home Costs Vary? The Determinants of Nursing Home Costs. *Howard Birnbaum, Christine Bishop, A. James Lee and Gail Jensen* 1095

Statistical Analysis of an Attitude Scale to Measure Patient Satisfaction With Medical Care. *Paula L. Stamps and Janice Berger Finkelstein* 1108

Life Satisfaction and Chronic Disease: Exploration of a Theoretical Model. *Julia S. Brown, May E. Rawlinson and Nancy C. Hilles* 1136

Refinements in the Measurement of Health-Specific Locus-of-Control Beliefs. *Richard R. Lau and John F. Ware, Jr.* 1147

BOOK REVIEWS

Managing Health Systems in Developing Areas, Experiences From Afganistan, edited by Ronald W. O'Connor. *Ray H. Elling* 1159

Health Plan, The Only Practical Solution to the Soaring Cost of Medical Care, by Alain C. Enthoven. *Rosemary Stevens* 1159

LETTERS TO THE EDITOR 1161

MEDICAL CARE
November 1981, Vol. XIX, No. 11

Original Articles

DRG-Based Reimbursement: The Use of Concurrent and Retrospective Clinical Data

CYNTHIA BARNARD, B.A.,* AND TRUMAN ESMOND, M.S.†

The Health Care Financing Administration is developing a new method of Medicare patient reimbursement limit setting, based on a quantification of hospital case mix through the AUTOGRP Diagnosis-Related Groups established at Yale University. The reliability of this method is dependent on the diagnosis and procedure data used as input, which should reflect the "principal" condition of the patient. HCFA's source of data is Medicare billing for 1978, some of which contains concurrent rather than retrospective diagnosis and procedure information. Billing data from one large teaching hospital are examined with respect to Medical Records data. The data are evaluated, based on the diagnosis and procedure codes and on the groupings (DRGs) presently being used by HCFA; concurrent and retrospective data are found to be widely divergent on both measures. An apparent difference in complexity or extent of resource use is noted, suggesting that the data being used in HCFA's development effort may not fully represent the level of complexity of cases being treated and that reimbursement based on this data may be incorrect.

IN RESPONSE TO THE REQUESTS of hospitals for a more appropriate system of reimbursement and to pressure from the federal government for a reduction in health care costs, the Health Care Financing Administration is in the process of developing a replacement for its current reimbursement mechanism for Medicare inpatient stays. This replacement has as its key attraction the establishment of reimbursement rates based on case mix, or the complexity of the cases treated in the institution over some period of time. It is intuitively obvious, say the proponents of this strategy, that costs are not uniform for the treatment of all diseases.

The methodology that is to be used to quantify case mix is one that was developed at Yale University and that is receiving a good deal of attention as the basis for New Jersey's present case-mix reimbursement experiment.[1] It was selected from a variety of case-mix measurement techniques that have been developed to aid in the evaluation of quality of care, resource needs, etc. Diagnosis-Related Groups (DRGs) are aggregates of patients based on five factors: age, principal and secondary diagnoses, principal and secondary procedures. They are claimed to be

---

* Senior Systems Analyst, Rush-Presbyterian-St. Luke's Medical Center, Chicago, Illinois.

† Vice President of Finance, Rush-Presbyterian-St. Luke's Medical Center, Chicago, Illinois.

Address for reprints: Cynthia Barnard, Data Center, 1725 West Harrison/11th Floor, Chicago, IL 60612.

0025-7079/81/1100/1071/$01.10 © J. B. Lippincott Co.

medically meaningful groups, with internally homogeneous patterns of length-of-stay.

The purpose of the present study is not the validation or discrediting of the DRG methodology per se, but rather a review of the appropriateness of the data utilized in its application. The following pages review evidence that the data being used by HCFA to develop DRG cost protocols and budgets may be of poor quality with respect to their present use, that is, not appropriate for DRG classification. These data consist of diagnosis and procedure information supplied on the final Medicare bill for an inpatient stay.

The areas of focus in this study are:

The ambiguity inherent in use of diagnosis and procedure coding schemes, and their replicability;

The source of Medicare bill data, and their relevance to DRG assignment;

Financial and case-mix implications of discrepancies between billing data and Medical Records data.

The Reliability of Coded Data

A variety of analyses have been performed to assess the validity of diagnosis and procedure assignment. These generally focus on the accuracy of the process that translates a physician's description of a disease or procedure into a standard code.‡ Estimated rates of error in this process vary from 5 per cent to as much as 34.8 per cent and higher.[2] The data supplied on the Medicare bill [form HCFA-1453(c); see Figure 1] are an English description; thus, HCFA is performing a coding process and runs the risk of error. The diagnosis and procedure descriptions are written on the bill according to no uniform standard; the physician's text is transcribed from the record. Corn[3] points out that 20 per cent of these descriptions are computer-coded into ICD codes§ and that the remainder are coded manually.

In 1977 a study was performed to audit ICD codes submitted on Medicare and Medicaid bills from 50 hospitals.[2] In only 66.8 per cent of the cases was the original diagnosis coding confirmed. The discrepancies between original coding and audit coding were usually due to "a difference of professional opinion in interpreting the medical record," suggesting that "for any aggregate sample of abstracts, a sizable portion will not be reliable in the sense of being coded similarly on repeated occasions. Furthermore, there is no identifiable or correctable 'error.'"[2]‖ A later study found the same situation.[1]#

Since Medicare does not use coded data from the hospitals, hospital-specific "error rates" cannot be determined. Medicare coders may or may not distribute "errors" across hospitals. If Medicare did request codes from hospitals, audits of each facility would be possible, intrafacility consistency would be more assured and constructive activity to achieve interfacility

---

‡ The coding scheme is called the International Classification of Diseases (adapted for use in the U.S.), or ICDA. In January 1979, Medicare stopped using the eighth revision of the system (ICDA-8) and began using the ninth, with "Clinical Modifications" (ICD-9-CM).

§ Julian Pettengill, from HCFA Office of Research, informs us that "the correct figure is about 50 per cent" for computer-coded diagnoses and procedures.

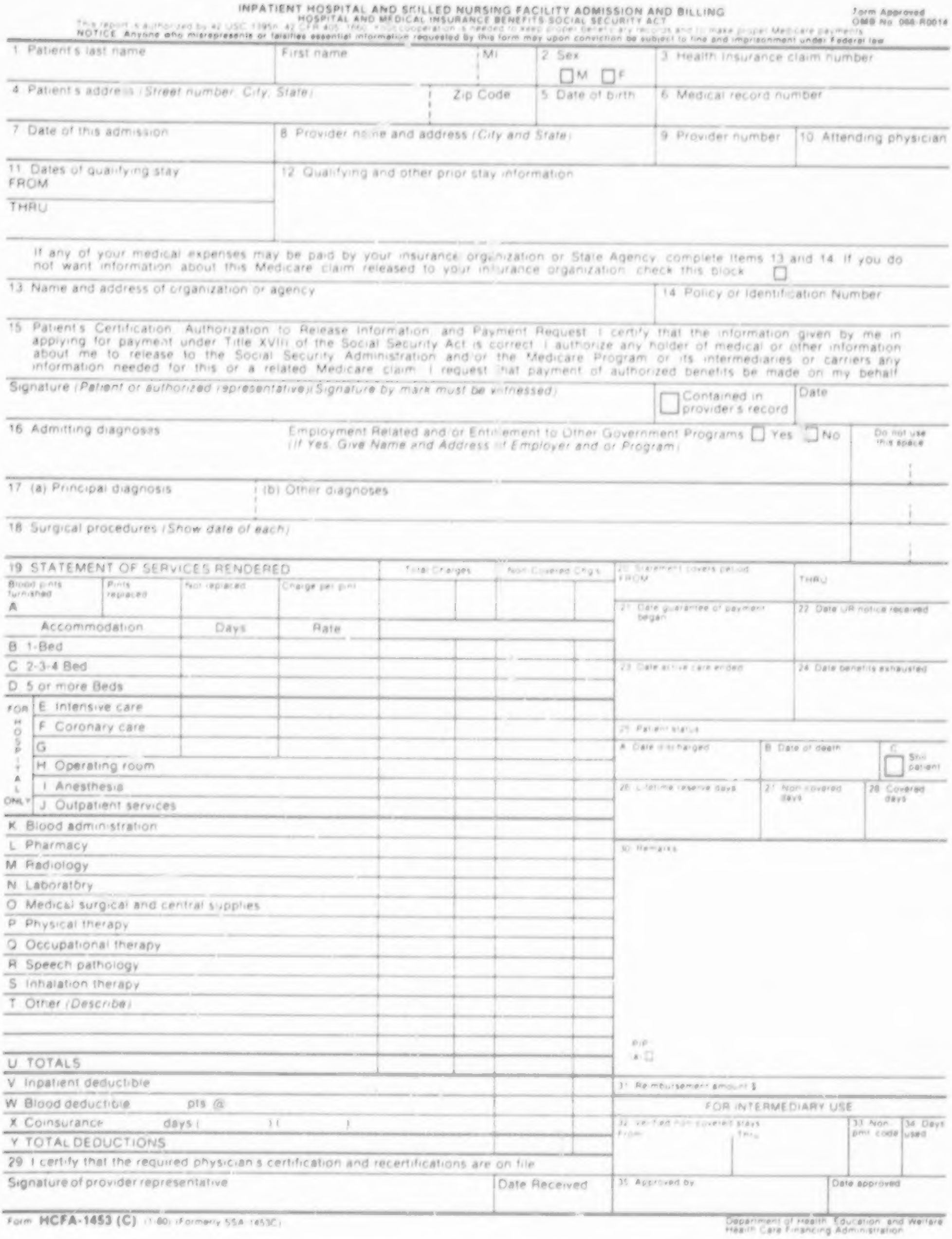
‖ Reasons for diagnosis code discrepancy:

professional opinion (on the code or the sequence of diagnoses):	47.6 per cent
specific hospital interpretation, procedure or policy:	42.2 per cent
other:	10.2 per cent

Reasons for procedure code discrepancy:

professional opinion:	35.1 per cent
policy:	53.1 per cent
other:	11.8 per cent

This study found a 39 per cent "error" rate on principal diagnosis (all four digits), of which 79 per cent was attributable to judgment or procedural differences.



INPATIENT HOSPITAL AND SKILLED NURSING FACILITY ADMISSION AND BILLING
HOSPITAL AND MEDICAL INSURANCE BENEFITS SOCIAL SECURITY ACT

Form Approved
OMB No. 068-R0014

This report is authorized by 42 USC 1395n, 42 CFR 405.1660. Your cooperation is needed to keep proper beneficiary records and to make proper Medicare payments.
NOTICE Anyone who misrepresents or falsifies essential information requested by this form may upon conviction be subject to fine and imprisonment under Federal law

1 Patient's last name | First name | MI
2 Sex ☐ M ☐ F
3 Health insurance claim number
4 Patient's address (*Street number, City, State*) | Zip Code
5 Date of birth
6 Medical record number
7 Date of this admission
8 Provider name and address (*City and State*)
9 Provider number
10 Attending physician
11 Dates of qualifying stay FROM THRU
12 Qualifying and other prior stay information

If any of your medical expenses may be paid by your insurance organization or State Agency, complete items 13 and 14. If you do not want information about this Medicare claim released to your insurance organization, check this block ☐

13 Name and address of organization or agency
14 Policy or Identification Number

15 Patient's Certification, Authorization to Release Information, and Payment Request. I certify that the information given by me in applying for payment under Title XVIII of the Social Security Act is correct. I authorize any holder of medical or other information about me to release to the Social Security Administration and/or the Medicare Program or its intermediaries or carriers any information needed for this or a related Medicare claim. I request that payment of authorized benefits be made on my behalf.

Signature (*Patient or authorized representative*) (*Signature by mark must be witnessed*)
☐ Contained in provider's record
Date

16 Admitting diagnoses
Employment Related and/or Entitlement to Other Government Programs ☐ Yes ☐ No
(*If Yes, Give Name and Address of Employer and/or Program*)
Do not use this space

17 (a) Principal diagnosis | (b) Other diagnoses
18 Surgical procedures (*Show date of each*)

19 STATEMENT OF SERVICES RENDERED				Total Charges	Non-Covered Chg's
Blood pints furnished A	Pints replaced	Not replaced	Charge per pint		
Accommodation		Days	Rate		
B 1-Bed					
C 2-3-4 Bed					
D 5 or more Beds					
FOR HOSPITAL ONLY E Intensive care					
F Coronary care					
G					
H Operating room					
I Anesthesia					
J Outpatient services					
K Blood administration					
L Pharmacy					
M Radiology					
N Laboratory					
O Medical surgical and central supplies					
P Physical therapy					
Q Occupational therapy					
R Speech pathology					
S Inhalation therapy					
T Other (*Describe*)					
U TOTALS					
V Inpatient deductible					
W Blood deductible	pts @				
X Coinsurance	days (	) (	)		
Y TOTAL DEDUCTIONS					

20 Statement covers period FROM | THRU
21 Date guarantee of payment began
22 Date UR notice received
23 Date active care ended
24 Date benefits exhausted
25 Patient status
A Date discharged | B Date of death | C ☐ Still patient
26 Lifetime reserve days | 27 Non-covered days | 28 Covered days
30 Remarks
PIP (a) ☐
31 Reimbursement amount $

FOR INTERMEDIARY USE
32 Verified non-covered stays From | Thru
33 Non-pmt code
34 Days used
35 Approved by | Date approved

29 I certify that the required physician's certification and recertifications are on file
Signature of provider representative | Date Received

Form **HCFA-1453 (C)** (1-80) (Formerly SSA-1453C)

Department of Health, Education, and Welfare
Health Care Financing Administration

FIG. 1.

consistency would be more easily implemented.[¶]

The 1977 Institute of Medicine study also focused on the issue of problems in coding that might be unique to Medicare: "abstracts for Medicare patients tended to be less reliable . . . Medicare patients are usually older and tend to have multiple diagnoses; this may explain the generally lower reliability. . . ."[2]

The picture that emerges is one in which professional judgment plays a significant role in code assignment, in which there is as a result some variability and lack of standardization, and in which there is apparently a higher rate of error when a complex case, presenting multiple problems, is coded. It seems reasonable to question the use of such data for reimbursement; we will discuss below the extent to which DRG aggregation eliminates some of the data problems.

The Problem of Definition

The Uniform Hospital Discharge Data Set (UHDDS) was developed for the U.S. National Committee on Vital and Health Statistics.[5] This document contains definitions of diagnoses and procedures, and the meaning of the term "principal."

> Principal diagnosis is . . . the condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care. . . . The principal procedure is one that was performed for definitive treatment . . . [and] is that procedure most related to the principal diagnosis.

Possibly the most important element of this definition is contained in the words "after study." Principal diagnosis refers to the outcome of a *retrospective* review process, conducted after patient discharge and based on all available data about the stay.

There are several alternatives to principal diagnosis when describing a patient. The "primary" diagnosis (the similar term is to be deplored and is responsible for a good deal of confusion) is "the one diagnosis, but not necessarily the principal diagnosis, that describes the most important or significant condition of a patient in terms of its implications for his health, his medical care, and his use of the hospital."[6] It is not clear from the UHDDS whether such a diagnosis could be established *during* a hospitalization, or whether this is a retrospective determination like that of the "principal" diagnosis.

The UHDDS devotes a thoughtful discussion to yet another type of diagnosis, the "presenting diagnosis or problem," which is concurrently determined and represents, at the time it is recorded, "the highest level of certainty supported by available clinical information." This diagnosis is the outcome of concurrent (e.g., Utilization Review) study and is the basis for decisions concerning the treatment regimen.

It is instructive to compare this 1980 version of the UHDDS to an earlier (1973) version,[7] in which "the recent development of problem-oriented records" (concurrently recorded) is noted. The 1973 version also anticipates future concerns of diagnosis-based research and patient management:

> Of all the items in the data set, the clear identification of diagnoses was considered of the greatest importance . . . standardization is an essential requirement for achieving comparability of data. . . .

As one observer has noted, "the Medicare definition of the principal discharge diagnosis does not appear to conform to the UHDDS. Medicare requires the hospital to report the condition that was the primary reason for the patient's *hospitalization*, rather than the primary reason for the patient's *admission* . . . the Medicare definition is somewhat vague."[3]

This vagueness has never been a problem, since the diagnostic data on the Medicare bill have never been used in such a way as to impact providers. A DRG-based reimbursement strategy, however, will change all this. Clearly, Diagnosis-Related Groups must be based upon the diagnosis and procedure information, which is the best determinant of resource consumption. To date, uniform standards for bill completion have not been disseminated to providers.

[¶] HCFA has begun the conversion effort to accept coded data from hospitals.

The Problem of Timing

The critical issue appears to be one of *timing*. In order to perceive this problem clearly, it is helpful to recall that the legislation mandating utilization review (PL 92-603, 1972) has been widely criticized for its fuzziness and apparent concern for cost-containment above quality assurance: "From the beginning SSA avoided enforcement of a qualitative review process, [and] treated utilization review as an audit of claims for payment rather than of professional practice."[8]

The relevance of these perspectives is evident when one reflects that *the source for the diagnosis and procedure information on the Medicare bill is often Utilization Review*, and Utilization Review evaluators are concerned with *concurrent*, not *retrospective*, diagnoses.

Fiscal survival dictates that all third-party payors be billed as soon after discharge as is practical—that is, as soon as all late charges have been posted. At Rush-Presbyterian-St. Luke's Medical Center (RPSLMC), a 1200-bed academic medical center in Chicago, Medicare bills are produced on the fourth day after discharge. Staff in the billing office use the computerized Utilization Review system to enable them to put the clinical data on the bills, since the medical record is not always completed at this point. The "current or discharge" diagnoses and procedures requested on form 1453 are thus available to HCFA—but they are often "current" as of a few days before discharge.

Medicare's attention to clinical data has always been cursory, and it seems fair to suggest that no institution will delay receivables collection in order to validate unaudited codes. The legion problems of timely medical record completion are well known, and there has never been a reason to hold Medicare bills up to 45 days (the PSRO requirement) or longer, waiting for the discharge diagnosis to be coded.

The Medicare Periodic Interim Payment (PIP) program requires that 90 per cent of patients' final bills be submitted within 30 days of discharge. In fact, Medicare has recognized this operational difficulty: "The intermediary may authorize the hospital to use the 'working' diagnosis (instead of the 'final diagnosis') to speed the submission of bills."[9] In short, some hospitals have used *concurrent* diagnosis and procedural data; others have used *retrospective* data. There is no doubt that only the latter is unquestionably suitable for DRG classification.

At the study hospital, Rush-Presbyterian-St. Luke's Medical Center, the validity of the codes that are entered by Utilization Review appears to be equal to that of the codes entered in Medical Records (based on informal audits conducted periodically by the directors of both departments). However, the coding activity serves different purposes. In Utilization Review, the goal is to establish the primary current problem, so that the timeliness of discharge may be reviewed. In Medical Records, the goal is to code "the diagnosis of the condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care."[6]

The RPSLMC Study

The foregoing tends to suggest that the data supplied to HCFA via form 1453, the Medicare bill, may be inappropriate for use in defining a reimbursement methodology, since distributions based on the data may be skewed. Because of its historical lack of importance, it may not be audited or validated by most hospitals, and because of the need to prepare Medicare bills on a timely basis, the data submitted may have been concurrent, not retrospective, data.

Objectives of the Study

A study was conducted at RPSLMC to compare billing data (concurrently determined and submitted to HCFA four days after discharge) with Medical Records data (retrospectively determined after discharge). The goal was to evaluate the extent to which concurrently collected data could reliably be substituted for retrospective data. Since HCFA is using diagnostic data to set up reimbursement protocols, we were particularly interested in the impact on reimbursement of concurrent data submission. In other words, if diagnostic data are used to set reimbursement levels, and an alternate source of data is used, what is the resulting effect on those reimbursement levels?

Dimensions of the Study

The data used for the study were complete medical records and utilization review records for approximately 3,000 Medicare inpatients discharged from RPSLMC during the year beginning May 1, 1979. This represented a random 50 per cent sample of the year's Medicare inpatients.

The two sets of data were merged by patient number (a combination of medical record number and a unique number representing a particular stay). The resulting data are presented in Table 1.

The data were entered in the GROUP computer program, yielding a concurrent DRG and a retrospective DRG, based on the two sets of codes.**

Results: Comparing the Raw ICD Codes

Based on the assumption that Medical Records coding reflects the accurate retrospective or discharge data, an attempt was made to compare the Utilization Review data—which are used on the Medicare bill—with this "criterion" data, in order to determine whether there are differences and, if so, the impact of those differences on reimbursement. The comparison was limited to the "integer" portion of the ICD code (up to three digits), making it as generous as possible, on the theory that the specificity provided by the two decimal places would be irrelevant to DRGs.

The question asked appears to be a simple one: for how many patients was con-

TABLE 1.

Both	Utilization Review (concurrent, used for billing)	Medical Records (retrospective)
Patient number	up to 5 diagnoses	up to 8 diagnoses
Age	up to 5 procedures	up to 8 procedures

** In order to be grouped, the codes had to be converted from the ICD-9-CM scheme to HICDA-2. No computer software—or manual methodology—has yet been developed to derive DRGs from ICD-9-CM codes. (A project to do this is presently underway, with HCFA funding.) The New Jersey Department of Health faces this issue in their DRG-based reimbursement system and kindly supplied RPSLMC with a conversion tape listing the ICD-9-CM codes and equivalent HICDA-2 codes.

Two documents from the New Jersey Department of Health describe the conversion strategy (the user must write the software to implement the conversion tape) and the Grouper strategy. Both are titled "A Prospective Reimbursement System Based on Patient Case-Mix for New Jersey Hospitals 1976–1983"; one is subtitled "DRG Assignment Using ICD-9-CM Codes: 1980 GROUPER Logic," and the other is subtitled "GROUPER INSTALLATION (for Modified HICDA-2 GROUPER)." Both documents are dated November 15, 1979 and are published by the Department of Health.

current coding the same as retrospective coding?

Our input consisted of two diagnoses and two procedures (one of each designated "principal" by the coder). For each of these four codes, then, we compared concurrent coding with retrospective coding. However, we recognize several kinds of "same" coding. If the principal retrospective diagnosis was coded concurrently as secondary, this was a separate category; if it was coded as third, fourth or fifth, this was separate. The reason for categorizing concurrent/retrospective coding similarities in this way has to do with DRG aggregations (see below). Table 2 summarizes the results.

Almost all cases (97 per cent) had a principal diagnosis assigned by Medical Records (retrospectively). In 35 per cent of the cases, the principal retrospective diagnosis was coded principal in concurrent Utilization Review as well; for 6.3 per cent, the retrospective code was matched, not by a principal, but by a secondary concurrent code; and for 2.8 per cent, it was matched by a tertiary code. In 53 per cent of the cases, however, the retrospectively coded diagnosis had not been coded *at all* on a concurrent basis.

Similar results were found when comparing principal procedure in concurrent and retrospective coding, although 15 per cent of the sample did not have a principal procedure assigned retrospectively by Medical Records. For 31 per cent of the cases, concurrent and retrospective coding agreed on principal procedure; and for 7 per cent, the concurrent "secondary" matched the retrospective "principal." for 47 per cent, the concurrent review did not yield the same code at all as that which would be coded by Medical Records after discharge.

For the limited number of cases for which a secondary diagnosis or procedure was assigned retrospectively, concurrent review usually did not yield the code at all. There was no secondary diagnosis for 41.6 per cent of the cases, and no secondary procedure for 40.19 per cent; the concurrent review failed to match the retrospective one that *was* assigned in 44.76 per cent of the cases on secondary diagnosis and 47.11 per cent on secondary procedure.

There is no suggestion here that concurrent coding is in error. On the contrary, by definition, concurrent review yields "the highest level of certainty supported by available clinical information." The last

TABLE 2. Matching Retrospective Principal and Secondary Codes to Concurrent Codes

	Principal Diagnosis	Secondary Diagnosis	Principal Procedure	Secondary Procedure
Retrospective code matches concurrent (%)	35.05	4.77	31.32	5.34
Retrospective principal as concurrent secondary or vice versa (%)	6.38	6.45	4.43	6.96
Retrospective code as concurrent tertiary (%)	2.82	2.41	.54	.40
No retrospective code found (%)	2.62	41.60	15.83	40.19
Retrospective code not in concurrent (%)	53.12	44.76	47.88	47.11
Total	100.00	100.00	100.00	100.00

Rounded numbers may not add to 100%.

concurrent review of a stay usually occurs several days before discharge; any clinical information that becomes available during those last few days cannot be reflected in the concurrent review records.

To return to the original question, then, it appears that in only about one third of the cases are concurrent and retrospective coding the same on any one diagnosis or procedure code. In some additional cases, they are "similar"—that is, the retrospective code was picked up concurrently, but it was listed as secondary instead of principal, tertiary instead of secondary, etc.

We still have not answered our original question about same overall coding; in order to do this, we must construct a four-dimensional matrix to show the interaction of sameness, similarity and difference on all four variables (two diagnoses, two procedures). For example, if a patient had the same principal diagnosis coded concurrently and retrospectively, then was the assigned principal procedure also the same? What about the secondaries?

This matrix was constructed, but only one portion of it is of interest. The DRG classification that HCFA is using differs from the original one in that principal diagnoses and procedures are used to group precisely, but secondaries are only "present" or "absent." That is, there is no discrimination based on "major" secondary codes.

Therefore, we are interested in that area of the four-dimensional matrix in which principal codes are the same in concurrent and retrospective coding and in which secondary codes are present or absent in both data sources. Only 292 cases, or 9.8 per cent, were the same in this elastic sense; that is, in only 9.8 per cent of the cases were principal diagnosis *and* principal procedure the same in concurrent and retrospective coding, while the secondary diagnoses and procedures were present or absent in both coding methods.

TABLE 3. Discrepancies in Retrospective Versus Concurrent DRGs

	No. of Cases	%
Retrospective DRG same as concurrent	684	22.984
Retrospective DRG differs from concurrent	2292	77.016

Implications for DRGs

When carrying the coding differences through the DRG aggregation process, we can expect *more* than 9.8 per cent of the cases to have the same DRG based on concurrent and retrospective coding, because each DRG is an aggregation of a number of diagnosis and procedure codes, and a certain amount of difference in the codes from the two sources will be obliterated by the grouping process.

This less sensitive measure, however, still reflected the discrepant natures of concurrent and retrospective coding, in that only 22.98 per cent of the cases remained in the same DRG (see Table 3). In general, examination of the data indicated that the major reason for a different DRG, from one data source to another, was a different principal diagnosis (as would be expected). Difference in a principal procedure was the next most likely cause. However, 17.31 per cent of all cases had different concurrent/retrospective DRGs, although the principal diagnosis and principal procedure were the same, suggesting that the secondary codes are not negligible. One hundred twenty-six of the DRGs are constructed with a simple distinction between presence and absence of a secondary code; for these, HCFA's methodology may be adequate. However, 39 additional DRGs include a distinction based on "major" versus "minor" secondary codes. The RPSLMC data support the hypothesis that accurate secondary coding is necessary for accurate DRG aggregation. This may be a problem for the HCFA simplification of the secondary codes in the modified grouping methodology.

Results: The Magnitude and Implications of DRG Differences from Concurrent and Retrospective Coding

A measure of the magnitude of a change in DRG was constructed from data provided by HCFA, which shows the "weight" assigned to each DRG, based on a preliminary analysis of a 20 per cent sample of 1978 Medicare bills. These weights reflect a manipulation of "average cost per case" based on Medicare cost-to-charge ratios and will presumably be used in calculating the overall "hospital case mix index" as described by Bentley and Butler.[1]†† The weights are expressed in dollar figures (e.g., $1195.77 for DRG 002) and represent calculated cost per case. It is important to note that this cost is based on HCFA's sample and that it is *not* necessarily reflective of resource use at the study hospital. However, it *is* based on a limited group of peer hospitals: the "Group 4" large, urban, 685-or-more bed institutions (regardless of teaching status).

Nine DRGs do not have any weight assigned because they were inadequately represented in HCFA's sample. A total of 28 concurrent DRG assignments and 21 retrospective assignments were made to these "unweighted" DRGs (a very small percentage of the sample). In the present study, zero was used as the default weight for these cases.

It is recognized that the use of weights is problematic, since it is based on the assumption that relative average cost equals relative actual case complexity, but there exists *no other* relative weighting of the DRGs. Furthermore, with the exception of these nine DRG weights, these numbers represent probable relative (if not actual) reimbursement and are the only figures that can be used to evaluate financial impact of the new system on an institution. (HFCA has not yet released the formula that will be used to calculate reimbursement amounts. Interim reports[13,14] suggest that case mix will be used as a weighting factor along with a local wage index, teaching factor, bed size and urban/rural index. Case mix has received the highest single weight in the equations thus far developed.)

Thus, the measure of the DRG charges was a simple arithmetic difference between the retrospective DRG weight and the concurrent one. For all cases, the average concurrent derived DRG weight was $2,611. The average retrospective-derived DRG weight was $3,212.

The average net difference was $601. The shift in mean weight between concurrent and retrospective coding is significant at the 0.0001 level (t-test under equal and unequal variance hypotheses).

For this sample, then, reimbursement based on concurrent data, with case mix and local wage index as the sole determinants of payment amount, would have averaged $600 less than reimbursement based on retrospective data. For an average annual load of 7,500 Medicare patients, this represents a total loss of $4.5 million. Of course, this is a "worst case" figure, since, in fact, teaching status and other factors will affect actual reimbursement amounts; still, the question of data collection time, of concurrent versus retrospective data, is far from academic when DRGs form a basis for a reimbursement methodology.

The Reasons for the DRG Differences

An analysis of variance was performed, using the four codes' "match" status as predictor variables and the net weight difference as the dependent variable.

---

†† For potential objections regarding the homogeneity of resource use within DRGs, note that the contention of the researchers who developed the Diagnosis-Related Groups is that the groups permit "the classification of patients according to their patterns of resource consumption while at the same time assuring that such classes [are] medically meaningful."[10,11]

Match status of the primary diagnosis would appear to be one major factor in explaining the net difference between concurrent and retrospective DRG. This is reasonable, since primary diagnosis is always the first-level aggregation criterion.[10,11]

An even more interesting result is the confirmation of the expected effect of the presence of procedures (especially surgical) on cost: the major factor, of the four considered, that explained the variation in net difference between retrospective and concurrent codes' "weights" was match status of the *primary procedure*.

Coding Audit

This study appeared to demonstrate substantial differences in the nature and purpose of coding clinical data between concurrent and retrospective coders. However, an alternate hypothesis might have been that one of the two coding departments had a high rate of error, yielding major differences in the resulting codes. In order to test this hypothesis, an audit was undertaken of a random sample of 100 cases in which the DRG derived from concurrent data differed from that derived from retrospective data.

Two senior coders, one from each department (Utilization Review and Medical Records), were given the patient charts and the computerized concurrent and retrospective codes. They were asked to determine why a discrepancy existed and, if an error had been made, to determine which code was correct. Some cases had more than one identifiable reason for the discrepancy; the 100 cases were assigned a total of 165 reasons. These reasons are shown in Table 4.

The cases were also analyzed by selecting the "most serious" of the reasons assigned to each, so that each case had only one reason associated with it. Most serious was defined as "closest to actual error" (see Table 5). In six cases, incorrect codes had been chosen by one or the other department; in 21 more, a clerical (keying) error meant that a code correctly transcribed on a record was incorrectly entered into the computer.

Sixty-eight of the remaining differences were determined to have been due to the correct performance of different duties in the two departments: timing differences meant that data were not always available to Utilization Review, and procedural differences meant that Utilization Review was deliberately focusing on current problems and complications rather than reviewing the entire stay. The last five cases

TABLE 4. Audit of Concurrent and Retrospective Coding Differences

Reasons for Code Differences (each case may have more than one reason assigned)	Frequency	Per Cent
Incorrect code selected by UR or MR	6	3.363
Professional judgment difference	20	12.121
Procedural difference in focus	72	43.636
Timing: data not available to UR	40	24.242
Clerical error in data entry to computer	25	15.152
Other	2	1.212
	165	100.000

UR = Utilization Review department.
MR = Medical Records department.

were simply "judgment calls," in which there was some room for professional difference of opinion about the precise code to be selected.

Overall, this audit appears to refute the hypothesis that either coding department was in error, but to support the hypothesis that concurrent clinical data reflect a coding process different from that of retrospective data and that the two may differ without either one being in error.

TABLE 5. Audit of Concurrent and Retrospective Coding Differences

Most Serious Reason for Code Differences (each case has only one reason assigned)	Frequency
Incorrect code selected by UR or MR	6
Clerical error in data entry to computer	21
Timing: data not available to UR	34
Procedural difference in focus	34
Professional judgment difference	5
	100

UR = Utilization Review department.
MR = Medical Records department.

Conclusions

The data upon which the Health Care Financing Administration plans to construct its new reimbursement mechanism may be inappropriate for that use. In how many hospitals were those data collected concurrently instead of retrospectively? In their concern for timely bill submission,‡‡ Medicare officials and hospital administrators may have unwittingly elected to sacrifice "final" diagnostic and procedural data. The concurrent data provided a problem indication that was sufficient for the reimbursement purpose of the time; it is, however, *not* sufficient for its proposed new purpose. Any provider who submitted concurrent data in past years may have distorted the case mix and runs a risk of incorrect—possibly inadequate—reimbursement under the new regulations.

Until the data can be adequately reviewed and explained, or replaced with appropriate retrospective data, and the differences between billing and clinical data eliminated, the current case mix based on Diagnosis-Related Groups cannot be used to measure resource use and therefore will not accurately predict reimbursement needs.

As Corn so astutely observes, "the producers of discharge data are not the principal users of the data . . . it appears that it is up to the data users to focus attention on the production of discharge data and to encourage uniform controls of their quality."[3]

In this case, the major "data user" is HCFA; the agency proposes to "focus attention" on discharge data by implementing a reimbursement system based on case mix. It seems evident that hospitals that supplied concurrent data during 1978 must furnish HCFA with retrospective data for that year, in order to establish a reliable data base of DRGs and associated costs. (It should be noted that the JCAH-mandated "Disease and Operations Index," maintained in each hospital, is an easily accessible source of this data.) Old data, however, are undoubtedly error-prone, as described in the Institute of Medicine study; Medicare providers and Medicare regulators will have to work together in coming years to improve the quality of the data before reliable case-mix reimbursement can be implemented.

---

‡‡ ". . . in 1979 . . . Medicare days at 49.3% . . . represent the highest number of revenue days outstanding for December 31 of any year since the survey began. . . . All payors registered increased days revenue in receivables over the same quarter last year . . . the 'Grand Total' for the quarter is 67.8% inpatient receivable revenue days outstanding." From: Chicago Hospital Council Memorandum No. 77, "Inpatient Receivable Survey Report—December 31, 1979." Chicago Hospital Council, April, 10, 1980.

References

1. Bentley, Butler. Case mix reimbursement: measures, applications, experiments. Hospital Financial Management 34:24, March 1980.

2. Institute of Medicine. Reliability of hospital discharge abstracts. Washington, DC: National Academy of Sciences, 1977.

3. Corn. Quality control of hospital discharge data. Med Care 1980;18:416.

4. Institute of Medicine. Reliability of national hospital discharge survey data. Washington, DC: National Academy of Sciences, 1980.

5. National Center for Health Statistics. Uniform hospital discharge data; minimum data set. Hyattsville, MD: DHEW, 1980. (Publication number (PHS) 80–1157.)

6. National Center for Health Statistics. Glossary of hospital terms. Chicago: American Medical Record Association, 1976.

7. National Center for Health Statistics. Uniform hospital discharge data: minimum data set. Hyattsville, MD: DHEW, 1973. (Publication number (HMS)73-1451.)

8. Feder M. Medicare: the politics of federal hospital insurance. Lexington, MA: Lexington Books, D.C. Health and Co., 1977.

9. Health Care Financing Administration, DHEW. Medicare hospital manual. Washington, DC: HCFA, 1980. (Publication number HIM-13.)

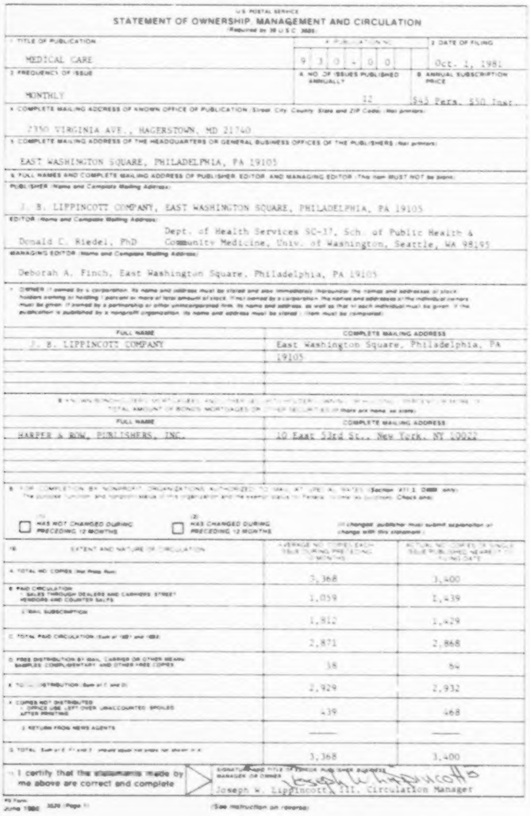
10. Mills, Fetter, Riedel, Averill. AUTOGRP: an interactive computer system for the analysis of health care data. Med Care 1976;14:603.

11. Fetter, Shin, Freeman, Averill, Thompsen. Case mix definition by diagnosis-related groups. Med Care (supplement), vol 18, Feb. 1980.

12. Young, Swinkola, Hutton. Assessment of the AUTOGRP patient classification system. Med Care 1980;18:228.

13. Pettengill, Vertrees. New uses for old data: a Medicare hospital case mix index. Health Care Financing Administration.

14. Lave, Pettengill, Vertrees, Lasdon. The Medicare hospital case-mix index. Health Care Financing Administration.



U.S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION: MEDICAL CARE

A. PUBLICATION NO.: 9 3 0 4 0 0

2. DATE OF FILING: Oct. 1, 1981

3. FREQUENCY OF ISSUE: MONTHLY

A. NO. OF ISSUES PUBLISHED ANNUALLY: 12

B. ANNUAL SUBSCRIPTION PRICE: $45 Pers. $50 Inst.

4. COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION (Street, City, County, State and ZIP Code) (Not printers): 2350 VIRGINIA AVE., HAGERSTOWN, MD 21740

5. COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS (Not printers): EAST WASHINGTON SQUARE, PHILADELPHIA, PA 19105

6. FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR, AND MANAGING EDITOR (This item MUST NOT be blank)

PUBLISHER (Name and Complete Mailing Address): J. B. LIPPINCOTT COMPANY, EAST WASHINGTON SQUARE, PHILADELPHIA, PA 19105

EDITOR (Name and Complete Mailing Address): Donald C. Riedel, PhD, Dept. of Health Services SC-37, Sch. of Public Health & Community Medicine, Univ. of Washington, Seattle, WA 98195

MANAGING EDITOR (Name and Complete Mailing Address): Deborah A. Finch, East Washington Square, Philadelphia, PA 19105

7. OWNER [illegible]

FULL NAME	COMPLETE MAILING ADDRESS
J. B. LIPPINCOTT COMPANY	East Washington Square, Philadelphia, PA 19105

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES (If there are none, so state)

FULL NAME	COMPLETE MAILING ADDRESS
HARPER & ROW, PUBLISHERS, INC.	10 East 53rd St., New York, NY 10022

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES (Section 411.3, DMM only) [illegible] Check one

(1) ☐ HAS NOT CHANGED DURING PRECEDING 12 MONTHS

(2) ☐ HAS CHANGED DURING PRECEDING 12 MONTHS

(If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES (Net Press Run)	3,368	3,400
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	1,059	1,439
2. MAIL SUBSCRIPTION	1,812	1,429
C. TOTAL PAID CIRCULATION (Sum of 10B1 and 10B2)	2,871	2,868
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	58	64
E. TOTAL DISTRIBUTION (Sum of C and D)	2,929	2,932
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	439	468
2. RETURN FROM NEWS AGENTS	——	——
G. TOTAL (Sum of E, F1 and 2 – should equal net press run shown in A)	3,368	3,400

11. I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Joseph W. Lippincott

Joseph W. Lippincott, III, Circulation Manager

PS Form 3526 (Page 1) June 1980

(See instruction on reverse)

MEDICAL CARE
November 1981, Vol. XIX, No. 11

A Case-Mix Method for Developing Health Planning Criteria for Hospital Services

JAMES GREENBERG, PH.D.,* AND ROGER KROPF, PH.D.†

The principal concern of this article is the inadequacy of the planning criteria currently being used by existing federally funded health planning agencies. A case-mix method was created for developing appropriate criteria. Chart-abstract data from New Jersey were used to create a list of those diagnoses eligible for treatment in the cardiac care unit (CCU), select a sample of hospitals for study and analyze the relationship between CCU bed need and CCU clinical practice. It was shown that existing bed-need criteria for CCU planning represent current CCU clinical practice patterns, which are probably not cost effective. The method was also used to develop empirical values of these criteria, which do represent cost-effective practice. It is recommended that the method be used to strengthen and update continuously all hospital-service need criteria used in current planning activities, including Plan Development and Certificate of Need Review.

RISING CONCERN over the escalating cost of health care[1-5] led Congress in 1974 to require that the Secretary of the Department of Health, Education and Welfare prepare and issue by regulation "National Guidelines for Health Planning."[6] Since then, resource standards have been issued as part of the Guidelines for a limited number of health care services, including acute-care hospital beds and computerized tomographic (CT) scanners.[7-9]

These standards are inadequate for the planning and regulation that health systems agencies and state health planning and development agencies currently undertake. They do not, for instance, cover many available health care services, such as cardiac care units and skilled nursing facilities. In addition, some of the standards, including those for acute-care hospital beds, have not been shown to be based on efficacious, cost-effective medical practice.‡ Also, none of them provides enough information to allow planners to adjust standards to reflect factors, such as socioeconomic status, that may cause variations in the need for services at the local level.[10,11]

A recent Institute of Medicine report states that the development of guidelines for health planning and regulation has been hampered by inadequate data bases, limitations in existing health planning

---

* Assistant Professor, Department of Health Science, Brooklyn College, Brooklyn, New York.

† Assistant Professor, Division of Public Health, University of Massachusetts, Amherst, Massachusetts.

This article is based on research that was jointly funded by the National Center for Health Services Research and the Bureau of Health Planning in the Department of Health, Education and Welfare under grant no. 1 R03 HS03538-01 NSS.

‡ As used in this study, cost-effective treatment means the least expensive form of treatment which has been proven to be therapeutically effective for the clinical conditions under consideration.

0025-7079/81/1100/1083/$01.10 © J. B. Lippincott Co.

analytic techniques and research methods, and insufficient knowledge of the efficacy of many health care services.[10] There is, therefore, an urgent need for a strengthening of the quantitative, analytic infrastructure of the health planning system. Specifically, there is an urgent need for empirically validated need criteria and procedures for adjusting these criteria to match local conditions in all existing health care facilities.

The research described in this article represents an attempt to address some of the issues raised in the Institute of Medicine report. Specifically, the authors developed and employed a new, inexpensive, flexible method for using hospital chart abstracts to create and analyze empirical values of need criteria. This method can also be used for the development of procedures for adjusting need criteria so that they account for the effect of local conditions on bed need.

The method allows the planner to go beyond the Lee–Jones methodology of the 1930s, which based criteria for determining the need for health care services on physician-defined clinical norms.[12] It achieves this by allowing for the use of evidence based on clinical research and on existing clinical practice in the creation of need criteria. Hence, it offers planners the opportunity to create and update need criteria continuously in order to reflect accurately optimal, current clinical treatment, and to use these criteria to plan for the rational allocation of the nation's health care resources.

Methods

A detailed description of the work on which this article is based has been presented elsewhere.[13] A brief synopsis is offered here.

The DRG Data Base

The New Jersey CCU system was selected as the field laboratory for this study because the state's health department has developed a comprehensive, statewide, computerized, hospital chart-abstract data base (the Diagnosis Related Group—DRG—data base) that is admirably suited for the field testing conducted in this research.[13]

CCU Bed-need Criteria

On the basis of a review of 207 health systems plans and annual implementation plans from health systems agencies across the country and of several other relevant documents on CCU planning,[14–17] nine different CCU bed-need criteria, which are currently being used by health planners, were identified.[13] Of these, the five criteria listed in Table 1 were chosen for further study.

These five criteria were chosen on the basis that, taken together, they provide the planners with the ability to determine CCU bed need in all regions of the country, no matter how rudimentary the local planning data base. The other four were not selected because none of them improves planners' ability to determine CCU bed need above that provided by the five criteria that were selected.[13] One of the criteria that was not selected, for instance, was the number of CCU admissions per 1,000 service area population. In order to use this criterion, the planner would require data on CCU service area population. However, if these data were available, the planner could use the third criterion that was chosen—the number of CCU beds per 100,000 service area population (see Table 1).

List of CCU-Eligible Diagnoses

A fundamental principle underlying this research is that the need for CCU beds depends primarily on case mix, which we have defined through an examination of diagnosis. A patient with a complicated acute myocardial infarction, for instance, will need a CCU bed, whereas a patient with only a simple pelvic fracture will not. In order to use diagnosis as a basis for determining CCU bed need, we required a list of those diagnoses that would make patients eligible for treatment in a CCU. A group of practicing northern New Jersey cardiologists directed the creation of this list. The list is based on the experience of this group of cardiologists, a review of the CCU literature and data derived from the DRG data base.[13]

The resulting list of CCU-eligible diagnoses, which was expressed in both H-ICDA-2 and ICDA-8 codes§ because both of these coding systems were used in New Jersey in 1977, was divided into seven categories. These categories, which were considered distinct in terms of their clinical characteristics and also in terms of CCU resource utilization, are listed in Table 2.

Selection of Hospitals for Study

Nineteen New Jersey hospitals were then selected for use in the study. They were selected because they had separate CCUs, rather than CCUs combined with intensive care units (CCU–ICUs), or no CCUs, and because of their accuracy in reporting CCU treatment in their chart abstracts.[13] Measured in terms of eight hospital characteristics, the sample of 19 hospitals was very similar to the whole population of New Jersey acute-care hospitals (Table 3). The 19 hospitals reported 225,886 chart abstracts in 1977, of which 9,193 were coded as having been CCU treated and which contained diagnostic codes on our list of CCU-eligible diagnostic codes. In other words, the empirical values of the CCU bed-need criteria reported here are based on a sample of 9,193 chart abstracts.

TABLE 1. Five CCU Bed-Need Criteria Chosen for Study

Criterion	Range of Values Now Used By Planners
Optimal average length of stay in the CCU	(2 to 6 days)
Optimal ratio of all CCU admissions to myocardial infarction CCU admissions	(1.5 to 2.5)
Expected no. of myocardial infarction admissions per 1,000 population	*
Optimal no. of CCU beds per 100,000 population	(3 to 12)
Optimal no. of CCU beds per 100 general-care beds	*

* No values for these criteria were found in our review of health planning documents.[13]

Measures of Current CCU Practice

Having thus defined a list of CCU-eligible diagnoses and selected the 19 hospitals for study, we proceeded to use

TABLE 2. Categories of Diagnoses Treatable in the CCU

DICAT No.	Diagnoses in DICAT*
1	Simple acute myocardial infarction (AMI)
2	Complicated AMI
1 + 2	AMI (simple or complicated)
3	Intermediate syndrome
4	Rule-out MI
5	Cardiac arrhythmia problems
6	Intensive cardiac conditions treatable in a regular intensive care unit
7	Other CCU-eligible cardiac conditions not often seen in the CCU
1–7	All CCU-eligible diagnoses
8	Diagnoses which are not eligible for CCU treatment

* DICAT represents diagnostic category.

---

§ This list of H-ICDA-2 and ICDA-8 codes is available from the first author upon request.

TABLE 3. Comparison of the 19 Selected Hospitals With All New Jersey Acute-Care Hospitals (1977)

Characteristics	19 Selected Hospitals	All NJ Hospitals
% Hospitals with residency programs in internal medicine[*]	10.5%	21.7%
% hospitals with residency program in cardiology[*]	5.3%	8.5%
Average no. of maintained CCU beds[†]	8.7	8.8[‡]
CCU occupancy rate[§]	79.9%	73.1%[‡]
% hospitals that have separate CCUs[§]	100%	29.2%
Average no. of hospital beds[§]	288	256
% hospitals that are governmental[‖]	5.3%	4.7%
% hospitals that are proprietary[‖]	5.3%	5.7%

[*] Source: Directory of Accredited Residencies in Internal Medicine, Annual Report on Graduate Medical Education in the United States, 1977–1978, American Journal of Cardiology, A.M.A.

[†] Source: New Jersey Department of Health.

[‡] The value of this rate for all New Jersey hospitals is based only on hospitals with separate CCUs.

[§] Source: "New Jersey General Hospital Facility Complement and Patient Data," 1977 Annual Summary (revised), New Jersey Department of Health, June 1978.

[‖] Source: Health Facilities Evaluation List of "Licensed General Hospitals in New Jersey," New Jersey Department of Health, June 1978.

them to define two measures of current CCU practice that are important in the determination of CCU bed need: CCU admitting criteria and CCU discharge criteria.[13]

This was achieved by first using the chart abstracts from the 19 hospitals to calculate the percentage of patients with CCU-eligible diagnoses who were treated in the CCU in each of the 19 selected hospitals in 1977, by diagnostic category. These data are presented in Table 4 and show, for instance, that in hospital 3, 49.0 per cent of all the patients discharged with simple AMI (DICAT 1) were treated in the CCU and that 29.3 per cent of patients discharged with complicated AMI (DICAT 2) were treated in the CCU. This finding was true in most of the 19 hospitals (Table 4) and was contrary to our original expectation that a higher proportion of complicated than of simple AMI patients would have been treated in the CCU.

The average values of these percentages in the 19 hospitals were used as measures of the stringency or looseness of the CCU admitting practices in these hospitals. Hence, hospital 39, for example, which treated the smallest percentage (5.2 per cent) of all admitted patients with any CCU-eligible diagnosis (DICATs 1–7) in the CCU, was considered to have the most stringent CCU admitting practices in the group of 19 hospitals. Similarly, hospital 12, which treated the largest percentage (26.1 per cent) of its admitted patients with any CCU-eligible diagnosis (DICATs 1–7) in the CCU, was considered to have the loosest CCU admitting practices.

The average CCU lengths of stay of CCU patients treated in the 19 hospitals in 1977 were taken as measures of the CCU discharge practices in the sample of 19 hospitals (Table 5). Thus, hospital 14, for example, which exhibited the longest average length of stay in the CCU for patients with any CCU-eligible diagnosis (DICATs 1–7), was considered to have very loose CCU discharge practices. Similarly, hospi-

TABLE 4. Percentage of CCU-Treated Discharges in Each of the 19 Selected Hospitals by DICAT (1977)

Hosp No.*	% Discharges With CCU-eligible Diagnoses CCU Treated in the following DICATS								
	Dicat 1	Dicat 2	Dicat 1 + 2	Dicat 3	Dicat 4	Dicat 5	Dicat 6	Dicat 7	Dicat 1–7
3	49.0	29.3	40.9	25.0	4.4	12.3	11.2	1.0	10.2
12	84.3	70.6	80.2	63.7	22.2	21.3	11.9	2.6	26.1
14	64.1	54.9	59.7	45.7	11.1	16.5	8.0	1.5	16.8
16	45.9	49.3	47.0	27.0	5.8	13.8	7.9	1.7	11.4
20	83.3	63.5	73.4	50.0	11.5	12.0	7.8	.8	14.7
26	85.1	66.2	77.9	69.9	12.4	16.4	14.3	1.4	19.7
29	83.2	75.5	81.0	51.6	16.8	26.1	31.3	4.8	21.6
39	34.9	25.3	32.5	13.1	1.7	3.7	2.9	.7	5.2
47	79.0	86.7	81.3	53.2	11.9	21.1	6.7	1.0	20.6
48	82.7	86.0	83.4	69.0	11.3	25.7	22.9	4.8	20.4
51	61.0	40.8	54.7	39.3	11.2	15.1	4.8	1.3	14.1
56	63.0	50.0	58.7	46.2	10.5	18.5	9.1	3.4	19.5
57	59.7	79.6	64.3	61.3	7.8	29.8	18.4	2.0	18.8
58	77.1	76.5	76.9	77.8	25.6	34.3	15.6	2.9	23.5
78	85.2	84.1	92.4	72.0	10.3	26.1	22.1	1.9	18.2
80	73.3	75.6	74.4	29.0	5.7	14.5	12.7	.4	6.9
85	64.1	66.7	65.3	43.3	3.4	20.2	15.5	.8	12.0
92	73.1	91.9	81.4	61.7	10.8	41.9	24.7	2.3	24.6
112	43.8	56.8	47.7	27.0	7.1	12.8	17.1	2.3	13.9
Mean	68.0	64.7	67.0	48.7	10.6	20.1	13.9	2.0	16.8
SD	15.8	19.3	16.6	18.5	6.0	9.0	7.5	1.3	5.8
Coeff. of Var. (%)	23.2	29.8	24.8	38.0	56.6	44.8	54.0	65.0	34.5

* These numbers were used in our data processing.[13]

tal 78, which had the shortest average length of stay in the CCU for patients with any CCU-eligible diagnosis (DICATs 1–7), was considered to have very stringent CCU discharge practices.

Empirical Values of CCU Bed-need Criteria

These measures of CCU admitting and discharge practices in the 19 selected hospitals, together with our list of CCU-eligible diagnoses, were the basic elements of our case-mix definition of current CCU practice that was used to calculate empirical values for the five selected CCU bed-need criteria (Table 6). The empirical values in Table 6 are average values for the 19 hospitals in the study. The empirical value for the first criterion, 4.25 days, for instance, was taken as the average length of stay in the CCU for patients with any CCU-eligible diagnosis (DICATs 1–7—see Table 5). The value for the second criterion was based on counts of the number of discharges with AMI and counts of the number of CCU-treated discharges, with both counts determined using chart-abstract data. Calculation of the third and fourth criteria made use of the size of each hospital's CCU service area, which was determined by means of a CCU patient-origin study, also using chart-abstract data.[13] The last two criteria made use of values of the number of CCU beds needed in each hospital, which was based on the number of CCU patient days actually used by CCU-treated patients with CCU-eligible diagnoses who were discharged in 1977. This latter calculation was also performed with the use of chart-abstract data.[13]

TABLE 5. Average Length of Stay in the CCU (Days) in 19 Selected Hospitals by Diagnostic Category (1977)

Hosp. No.*	Dicat 1	Dicat 2	Dicat 1–2	Dicat 3	Dicat 4	Dicat 5	Dicat 6	Dicat 7	Dicat 1–7
3	4.77	7.48	5.58	4.82	5.77	4.91	4.23	7.00	5.25
12	4.82	5.78	5.07	3.20	2.87	3.08	3.06	2.57	3.60
14	10.22	10.09	10.16	8.67	7.41	8.39	7.36	6.55	8.65
16	4.50	7.35	5.48	3.34	3.69	4.75	3.85	3.14	4.38
20	8.10	9.00	8.49	4.09	3.46	4.01	2.85	2.33	5.36
26	4.44	3.87	4.26	3.31	3.09	3.37	3.34	2.66	3.58
29	3.00	†	3.00	2.00	2.50	2.66	12.00	†	3.75
39	5.87	7.60	6.19	3.89	3.51	4.04	4.33	2.80	5.00
47	5.33	5.82	5.49	3.65	3.20	3.61	5.20	1.00	4.14
48	4.00	5.81	4.41	2.62	2.05	2.89	2.38	2.04	2.87
51	4.80	6.16	5.11	3.64	3.15	3.30	2.95	2.70	3.85
56	4.77	7.61	5.56	4.55	3.67	3.77	4.50	3.28	4.63
57	4.15	5.89	4.64	2.93	2.77	3.50	3.31	2.66	3.40
58	3.59	5.84	4.32	3.57	2.71	4.70	3.22	1.50	3.48
78	3.66	2.00	3.25	†	3.11	1.81	4.00	1.00	2.75
80	5.80	5.96	5.88	4.34	3.39	3.59	4.55	1.00	4.43
85	6.20	5.00	5.63	3.67	3.61	3.92	5.18	1.00	4.35
92	4.73	5.77	5.25	2.99	2.89	4.00	3.83	2.33	3.64
112	4.36	6.12	4.98	2.78	2.71	3.41	3.03	2.00	3.54
Mean	5.11	6.28	5.40	3.75	3.45	3.83	4.38	2.98	4.25
SD	1.68	1.80	1.63	1.40	1.22	1.33	2.17	1.91	1.30
Coeff. of Var. (%)	32.9	28.7	30.2	37.3	35.4	34.7	49.5	64.1	30.6

* These hospital numbers were used in our data processing.[13]

† Values were not able to be calculated for these elements because of poor coding of length of stay in the CCU on the chart abstract.

Study of the bed-need criteria in the 19 hospitals revealed considerable hospital-to-hospital variation. For example, the hospital-to-hospital variation in the first criterion listed in Table 6 (average length of stay in the CCU) is illustrated in Table 5. This type of hospital-to-hospital variation in such bed-need criteria suggests hospital-to-hospital variation in such factors as: maintained CCU bed levels; CCU admitting and discharge criteria and treatment practices; CCU patient characteristics; and CCU service-area characteristics.

A statistical analysis of CCU utilization in the 19 hospitals helped explain these variations in the empirical values of the bed-need criteria. This analysis considered the effect of factors such as clinical severity, patients' ages and CCU occupancy rates. Incidentally, this statistical analysis was also used to illustrate the development of procedures for allowing the health planner to adjust CCU bed-need criteria to account for factors which cause a legitimate, abnormally high or low CCU bed need on the local health service area level. As a result, it was concluded that the case-mix method is well suited for the development of such adjustment procedures for need criteria for all hospital facilities. This has been reported elsewhere.[13]

Although there were considerable hospital-to-hospital variations in the empirical values of these criteria, there was close agreement between the average value based on our definition of current CCU practice and that based on the actual numbers of CCU beds maintained in all 19

TABLE 6. Empirical Values of CCU Bed-Need Criteria in 19 Selected Hospitals (1977)

Bed-Need Criterion	Values Based on our Definition of Current CCU Practice*		Values Used By Health Planners
	Mean	SD	
Average length of stay in CCU (days)	4.25	1.30	2 to 6
No. CCU Discharges per AMI discharge	2.58	.65	1.5 to 2.5
No. AMI discharges per 1,000 service-area population	2.26	.57	†
No. CCU beds per 100,000 service-area population	6.28	1.30	3 to 12
No. CCU beds per 100 general care beds	1.99	.53	†

* Current CCU practice is based on our list of CCU-eligible diagnoses and the CCU admission and discharge practices in the 19 selected hospitals.

† No values were found for these criteria in our search of planning documents.[13]

hospitals (see Table 7).[‖] This suggested that the number of CCU beds in the 19 hospitals, taken as a group, was close to need, based on our definition of current CCU practice in 1977. However, this was not necessarily true in individual hospitals in the group.

Empirical values of the CCU bed-need criteria were generally close to or in the range of values used by CCU planners (Table 6). However, the range of values used by planners is extremely broad (Table 6), broader than seems appropriate for planning criteria. Use of values at the upper end of the range could lead to unnecessarily high CCU bed levels, utilization and expenditures. Similarly, use of values at the lower end of the range could lead to unnecessarily low CCU bed levels and high cardiac mortality rates. This underlines the urgent need for a far narrower range of empirically based CCU bed-need criteria.

Cost-effective CCU Bed-need Criteria

In an attempt to fill this need and in order to develop cost-effective CCU bed-need criteria, we used the case-mix approach and chart-abstract data from the 19 selected hospitals to calculate the values of bed-need criteria representing cost-effective CCU care.[13] A cost-effective CCU was defined as one that admits only patients with diagnoses for which the unit is the least expensive proven mode of therapeutically effective treatment, and that uses therapeutically sound but stringent discharge criteria. The selection of diagnoses for which the CCU has not been rigorously proven to be cost-effective, those in the first, fifth, seventh and eighth

---

[‖] There are only two criteria in this table because only two of the five selected CCU bed-need criteria could be related to the number of maintained beds (see Table 1).

TABLE 7. Average Empirical Values of CCU Bed-Need Criteria in 19 Selected Hospitals Based on Our Definition of Current CCU Practice and on the Number of Maintained Beds in 1977

Bed-Need Criterion	Value Based on No. of Maintained CCU Beds	Values Based on Our Definition of Current CCU Practice*	
		90% Bed Availability†	Average Bed Need‡
No. of CCU beds per 100 general care beds	3.11	3.10	1.99
No. of CCU beds per 100,000 service-area population	10.20	9.85	6.28

* Current CCU practice is based on our list of CCU-eligible diagnoses and the CCU admission and discharge practices in the 19 selected hospitals.

† In order to be able to compare with the number of maintained CCU beds in each hospital, our case-mix estimate of average bed need was used to calculate bed need for 90% bed availability, using the Poisson formula.[13]

‡ These values are from Table 6.

diagnostic categories in Table 2, was based on our review of the CCU literature.[13]

The selection of stringent discharge criteria was achieved by selecting an appropriate but low average CCU length of stay. For this purpose, the average CCU length of stay in hospital 48 (see Table 5) was selected, based on an examination of the average length of stay in the CCUs in the 19 selected hospitals.[13] Hospital 48 was selected because it had the lowest average length of stay for CCU-eligible discharges (DICATs 1–7) of the hospitals that provided enough well-coded chart-abstract data to allow for the calculation of average lengths of stay for each of the seven CCU-eligible diagnostic categories (Table 5). Hospital 78, which had a lower average length of stay for CCU-eligible diagnoses than hospital 48, was not selected because it did not provide enough chart-abstract data to enable the calculation of values of average length of stay for all seven diagnostic categories considered in this study (see Table 5). Hence, it was felt that use of data from hospital 78 might not yield representative empirical values of CCU bed-need criteria.

The resulting values of the cost-containment-oriented CCU bed-need criteria representing cost-effective CCU care were at the lower end of the ranges of values found in our review of CCU planning documents and literature (Table 8). They were used as a basis for recommending values of the five selected CCU bed-need criteria (Table 8) for CCU planning. In the interest of cost containment it is recommended that planners use these bed-need criteria until rigorous clinical trials support the expansion of the use of the CCU beyond the level on which the recommended bed-need criteria are based. When used for planning on the local, area-wide level, these criteria should be adjusted to account for factors that effect CCU bed need on the local level.

Discussion

The use of chart abstracts and the case-mix method for determining empirical values of need criteria is a powerful, efficient approach with several disadvantages. Probably the most serious disadvantage is that this approach uses the chart abstract as

TABLE 8. Values of Five CCU Bed-Need Criteria

Bed-Need Criterion	Empirical Value Based on Cost-Effective Treatment	Value Recommended for Planning	Values Used by Health Planners
Average length of stay in CCU (days)	2.98	3	2 to 6
No. CCU discharges per AMI discharge	1.65	1.5	1.5 to 2.5
No. AMI discharges per 1,000 service-area population	0.92	1	*
No. CCU beds per 100,000 service-area population	2.82	3	3 to 12
No. CCU beds per 100 general care beds	0.90	1	*

* No values for these criteria were found in our review of health planning documents.[13]

a data source. Chart abstracts contain errors, some of which are frequent and serious.[18,19] These errors derive from the fact that the abstracting process introduces errors of omission, miscoding and misinterpretation, and superimposes them on the errors or lack of documentation existing in the patient's chart.

Due to the burgeoning use of chart abstracts for health care reimbursement and evaluation,[20] chart-abstracting errors have been rather extensively studied.[18,19,32] Table 9, for instance, contains some of the results of a 50-hospital reabstracting study conducted by the Health Research and Educational Trust, an affiliate of the New Jersey Hospital Association.[19] These results suggest that two of the abstracted items used in this study, the discharge and admission dates, were reliably coded—the rate of agreement between the original code and the reabstracted code for both items was 99.7 per cent. On the other hand, they also suggest that principal and secondary diagnoses, which were important variables in our study, were rather poorly coded—with rates of agreement of 72.6 per cent and 66.5 per cent for principal and secondary diagnoses, respectively.

While these findings might argue that diagnosis was poorly coded, it should be pointed out that the HRET abstracting study considered the order in which diagnosis was coded. This was not a critical concern in our research because we used all diagnoses on the chart abstract, without regard to their order on the abstract.

Another disadvantage that could be attributed to the case-mix method is that it is only usable in areas of the country in which chart-abstract data bases have already been created. This is because the expense of creating a chart-abstract data base large enough to provide reliable information for the purpose of using this method de novo would probably not be justified when compared with the lesser expense of using one of the other methods of creating need criteria, such as past utilization or direct examination of a sample of patients or their charts.

There are only about 10 states in which chart-abstract data bases are currently being collected for use in the public do-

TABLE 9. Rates of Agreement Between HRET and Original Abstracts

Data Item	Rate of Agreement (%)
Date of birth	99.3
Sex	98.6
Race	96.9
Zip code	98.7
Admission date	99.7
Discharge date	99.7
Patient disposition	98.7
Principal payment source	97.5
Principal diagnosis	72.6
Secondary diagnosis	66.5
Other diagnosis	52.9
Principal procedure	66.4
Secondary procedure	59.3
Other procedure	44.0

Source: "Reliability of New Jersey Hospital Discharge Abstracts: Final Report," Health Research and Educational Trust, Princeton, NJ, July 1978.

main, so that this method is only usable in a small portion of the U.S. This implies that the method is probably not now feasible for the creation of need criteria representing clinical practice in most of the country. However, chart-abstracting organizations such as the Professional Activities Study do collect and process chart abstracts for most hospitals in the other states, so that this method could conceivably be used in all areas of the country. This is not now possible because, as a rule, hospitals do not make chart-abstract data available for public use.

One of the most significant advantages of the case-mix method is that it is able to use very large samples of patients and providers at very little expense in areas which have publicly available chart-abstract data bases. The 1,109,286 1977 New Jersey hospital chart abstracts used in this study, for example, were obtained at a cost of a few hundred dollars. The reason for this is that the expense of creating the New Jersey hospital chart-abstract data base is borne by the state, the federal government and other third-party payers. Funds also derive from the state's federally funded prospective reimbursement demonstration project.[20]

Two of the most powerful features of this method are that it is based on case mix and that it is able to incorporate the results of clinical research and other empirical information. These features enabled the method to be used for the characterization, analysis and evaluation of the clinical norms currently being used by CCU cardiologists. Similarly, these features account for the method's ability to specify clinical norms representing what we consider to be cost-effective CCU care, and to determine bed-need criteria representing this cost-effective care.

This result seems important because all existing need criteria are based on subjective, physician-specified clinical criteria, which presents several problems. The first is that physicians often differ widely in their specification of "appropriate" clinical criteria. In addition, there is a considerable body of literature suggesting that physicians sometimes use clinical criteria that yield treatment practices that have not been proven to be efficacious or cost-effective.[21-28] As an example, the acute-care hospital bed-need criterion of 4 beds per 1,000 population,[7] issued by the Department of Health, Education and Welfare—now called Health and Human Services—has been criticized by many physicians as being too skimpy. However, studies of bed need in HMOs suggest that this criterion could safely be lowered to about 2.5 beds per 1,000 population.[21,29] It has been suggested that this difference of opinion can be accounted for by the differences in clinical criteria used by physicians in HMOs and in fee-for-service practice. This does not necessarily indicate clinically inappropriate practice on the part of either HMO or fee-for-service practitioners. It could simply be a reflection of the lack of validated, specific norms for the need for clinical treatment.

Conclusions and Recommendations

Use of chart abstracts and the case-mix method has yielded results that suggest that the CCU bed-need criteria now being used by health planners are generally representative of current CCU practice (Table 6), which, on the basis of available evidence, is probably not cost-effective. The range of values of CCU bed-need criteria currently being used by health planners is too broad to allow for the allocation of CCU facilities in a rational, cost-effective manner.

In an attempt to strengthen the methods available to planners, therefore, we used the case-mix method to develop CCU bed-need criteria that do represent cost-effective CCU care, and we recommend their use for cost-containment-oriented CCU planning. In addition, in the interest of reducing unnecessary expenditures on CCU care in a rational therapeutically valid manner, it is strongly recommended that a rigorous clinical evaluation of current CCU practice be undertaken.

The case-mix method underlying this study is a powerful, inexpensive tool for relating the clinical, cost and utilization aspects of hospital care, and it is recommended that it be used, together with the chart-abstract data bases currently being created in several states, to validate and extrapolate the conclusions of this study. This includes the validation of our empirical values of CCU bed-need criteria (Tables 6 and 8). It also includes the development of adjustment procedures for CCU bed-need criteria so that they are able to account for factors that do cause legitimate extraordinary CCU bed need on the local level. This should be based on a comprehensive analysis of such factors as service-area population socioeconomic and demographic characteristics, which cause variations in CCU bed need on the local level.

The case-mix method should also be used to evaluate the cost-effectiveness of other forms of hospital care and the planning criteria used in their need determination. Where necessary, empirical values of need criteria representing cost-effective treatment and adjustment procedures for use on the local level should be developed. This should be done for services such as the intensive care unit, intermediate cardiac care unit, general care and diagnostic services such as the CT scanner and the electronic fetal monitor. This is particularly important in this time of concern over the increasing cost of health care, which has led to an urgent need to allocate health care services in a cost-effective and therapeutically valid manner.

Implementation of these recommendations would provide the health planner with continuously updated, empirically based need criteria that do represent cost-effective and therapeutically effective clinical practice, and that also account for clinical, demographic, geographic and socioeconomic variables that have a valid effect on the need for health care services. These criteria would be suitable for use by existing planning and reimbursement agencies for certificate of need review, plan development, appropriateness review and rate setting.

Although chart-abstract data have become increasingly available to individual hospitals, such data have not been readily available to planning and regulatory agencies, or for research purposes. While the Social Security Administration, and now the Health Care Financing Administration (HCFA), have been receiving these data on a 20 per cent sample of Medicare patients,[30] most other chart abstracts have not been in the public domain. A few states, including New Jersey and New York, now require the submission of discharge abstracts on all patients as part of HCFA-funded demonstration projects.

The controversy concerning the costs and benefits of implementing a uniform hospital financial report[31,32] suggests that the creation of a nationwide system for collecting chart abstracts will depend on the ability of planners, regulators and researchers to document the need for these data. Hopefully, this paper will strengthen the case for collecting and disseminating chart-abstract data for purposes of understanding and rationally regulating and monitoring our hospital system.

References

1. Council on Wage and Price Stability. The complex puzzle of rising health care costs. Washington, D.C.: Executive Office of the President, Dec. 1976.

2. Zubkoff M, Raskin IE, Hanft RS, eds. Hospital cost containment. New York: Prodist, 1977.

3. Greenberg, JA. Health care cost containment—a survey. Rochelle Park, New Jersey: Bergen-Passaic Health Systems Agency, May 1978.

4. Council on Wage and Price Stability. The rapid rise of hospital costs. Washington, D.C.: Executive Office of the President, Jan. 1977.

5. Council on Wage and Price Stability. The problem of rising health care costs. Washington, D.C.: Executive Office of the President, April 1976.

6. P.L. 93-641, Sec. 1501.

7. Federal Register, 42. National guidelines for health planning. Sept. 23, 1977: 48501-505.

8. Danacenn P. The health planning guidelines controversy: a report from Iowa and Texas. Washington, D.C.: Health Resources Admin., 1979. (DHEW Pub. No. (HRA) 79-647.)

9. Zwick DI. Initial development of national guidelines for health planning. Public Health Reports, 93, Sept.-Oct. 1978.

10. Health planning in the United States: issues in guideline development. A Report of a Study by the Institute of Medicine of the National Academy of Sciences, Washington, D.C.: Institute of Medicine, March 1980.

11. Shonick W. Elements of planning for area-wide personal health services. St. Louis: CV Mosby, 1976.

12. Lee RJ, Jones LW. The fundamentals of good medical care. Chicago: Univ. of Chicago Press, 1933. (Publications on the Cost of Medical Care, No. 22.)

13. Greenberg J, Haft J, Budke G, Roth J. A diagnosis-based study of the planning of cardiac-care units. Washington, D.C.: Natl. Ctr. for Health Services Research, 1979. (Final report of research grant no. HS03538.)

14. A. D. Little Inc. Planning for coronary care units: a technical assistance document for planning agencies. U.S. Dept. of Health, Education & Welfare, January 1979. (Publ. no. (HRA) 79-14019.)

15. Inter-Society Commission for Heart Disease Resources. Resources for the optimal care of patients with acute myocardial infarction. Circulation XLIII, A-171 thru A-183, 1971.

16. World Health Organization, The organization of coronary care units. Copenhagen: 1969.

17. Clipson CW, Wehrer JJ. Planning for cardiac care: a guide for the planning and design of cardiac care facilities. Ann Arbor: Health Admin. Press, 1973.

18. Report of a study: reliability of Medicare hospital discharge records. Washington D.C.: Institute of Medicine of the National Academy of Sciences, November 1977.

19. Reliability of New Jersey hospital discharge abstracts. Princeton: Health Research and Educational Trust of New Jersey Final Report, July 1978.

20. A prospective reimbursement system based on patient case-mix for New Jersey hospitals 1976-1978. Annual Report of the Diagnosis Related Group Project funded under Contract No. 600-77-0022. Trenton: Department of Health, Vol. 1, November 20, 1977.

21. McClure W. Reducing excess hospital capacity. Excelsior, Ma.: Interstudy, October 1976.

22. Roemer, I. Bed supply and hospital utilization: a natural experiment. Hospitals 1960; 36:36.

23. Bunker P. Surgical manpower: a comparison of operations and surgeons in the United States and in England and Wales. N Engl J Med 1970; 282:135.

24. Fuchs VR. Who shall live? Health economics and social choice. New York: Basic Books, 1974.

25. Medical technology and the health care system: a study of the diffusion of equipment-embodied technology. A Report of the Committee on Technology and Health Care, Assembly of Engineering, National Research Council and Institute of Medicine. Washington, D.C.: National Academy of Sciences, 1979.

26. Assessing the efficacy and safety of medical technologies. Washington D.C.: Office of Technology Assessment, Congress of the United States, Sept. 1978.

27. Roemer MI, Shonick W. HMO performance: the recent evidence. Milbank Mem Fund Q 1973; 51:271.

28. Somers AR, ed. The Kaiser-Permanente medical care program: a symposium. New York: Commonwealth Fund, 1971.

29. Ensminger B. The $8 billion hospital bed overrun: a consumer's guide to stopping wasteful construction. Washington, D.C.: Public Citizen's Health Research Group, 2000 P St. N.W., 1975.

30. Office of Program Policy and Planning, Social Security Administration. Medicare statistical files manual. ORS Pub. No. (7-77), p. 81.

31. Korus B. Uniform reporting in hospitals: What has been done. Topics in Health Care Financing 1979; 6:18.

32. Reliability of national hospital discharge survey data. Washington, D.C.: Institute of Medicine of the National Academy of Sciences, 1980. (IOM pub. No 80-02.)

MEDICAL CARE
November 1981, Vol. XIX, No. 11

Why Do Nursing Home Costs Vary?

The Determinants of Nursing Home Costs

HOWARD BIRNBAUM, PH.D.,* CHRISTINE BISHOP, PH.D.,† A. JAMES LEE, PH.D.,‡ AND GAIL JENSEN, M.S.§

Since the costs of nursing home care are a major component of the rapidly rising costs of health care, it is appropriate to base public policy discussions about cost containment on the determinants of nursing home costs. This article investigates the determinants of nursing home operating costs and reviews the results of 11 related econometric cost analyses conducted by the authors. Single-equation cost analyses are developed for nursing homes in three states and in the nation. The cost results of a multi-equation model of nursing home behavior are also reviewed. The analyses indicate that facility size and occupancy rate are minimally important in determing cost variation. Facility characteristics, particularly type of facility and ownership, are important variables. Nonprofit facilities consistently had higher costs than for-profit facilities, after controlling for patient mix and service differences, and, in one analysis, for a measure of quality.

THIS ARTICLE REVIEWS and integrates the results of 11 distinct econometric cost analyses conducted by the authors on 20 sets of nursing home data.[||] The data include information on costs, services produced, and characteristics of residents and facilities. The analyses include:

- a single-equation cost model for a national sample of 1,127 nursing homes, using the 1973-74 National Nursing Home Survey, from a survey conducted by the National Center on Health Statistics;
- a multi-equation, two-stage least squares (2SLS) model of nursing home behavior that includes a cost equation as well as equations for services provided, occu-

---

* Abt Associates Inc., Cambridge, Massachusetts.

† Center for Health Policy Analysis and Research, Brandeis University, Waltham, Massachusetts.

‡ Arthur D. Little, Cambridge, Massachusetts.

§ Center for Health Services Research and Department of Economics, University of Minnesota, Minneapolis, Minnesota.

This paper was prepared as part of a larger study of nursing home cost and reimbursement conducted for the Health Care Financing Administration under Contract No. 600-77-0068. The contents of this article do not necessarily reflect the views or policy of the Department of Health and Human Services or the Health Care Financing Administration.

The authors wish to thank Dr. Thomas Kickham, the HCFA Project Officer.

Address for reprints: Howard Birnbaum, Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138.

[||] The approach used in this paper to synthesize results across the multiple data sources is similar to that used in a related review of some of the nursing home cost literature in Bishop.[1] An interpretation of the results described in this paper for use in improving nursing home reimbursement policy is contained in Birnbaum et al.[2]

0025-7079/81/1100/1095/$01.15 © J. B. Lippincott Co.

pancy rates, the mix of public and private patients, and the rate charged to private patients. These analyses were conducted using the national sample of nursing homes mentioned above;

- three single-equation cost models of nursing homes in New York, using the Medicaid Cost Reports for facilities in that state. Separate cost analyses were conducted on approximately 500 facilities for each of three years: 1974, 1975 and 1976;
- three single-equation cost models of nursing homes in Massachusetts, using the Medicaid Cost Reports for facilities in that state. Separate cost analyses were conducted on approximately 450 facilities for each of three years: 1973, 1974 and 1975;
- three single-equation cost models of nursing homes in Indiana, using Medicaid data for facilities in that state. Separate cost analyses were conducted on approximately 80 facilities for each of three years: 1973, 1974 and 1975.

The econometric studies used distinct sets of nursing home data, for different times and for different geographic areas (Table 1). The "cost function" analysis has identified variables that are statistically significant in explaining variation across providers and in assessing the magnitude and direction of their effects. The cost function of economic theory is a measure of the relationship between average cost per unit of output and rate of production However, the cost analyses described here also allow average cost to vary systematically by provider type, input price level and output.

The various econometric analyses were reasonably successful in accounting for the variation in the dependent average operating cost variable. Across the data sets, the explanatory variables explain or predict between 59 and 77 per cent of the variation. The studies also show that many variables are strongly and consistently associated with operating costs. However, it can be misleading to compare actual regression coefficient values across the studies, since the other variables "held constant" by regression analysis differ. Also, the time periods differ.

Average operating cost is measured as per diem total cost net of reported capital-related expense.# For a variety of reasons, it

TABLE 1. Description of Nursing Home Samples

Geographic Location	Year	No. of Providers	R^2*	Data Source
Nationwide	1972	1127	.585	National Center for Health Statistics
Massachusetts	1973	453	.656	Massachusetts Dept. of Public Health
	1974	468	.713	Rate Setting Commission
	1975	438	.719	
New York	1974	479	.647	New York State Dept. of Health
	1975	504	.768	
	1976	504	.764	
Indiana	1973	78	.721	Indiana Dept. of Public Welfare
	1974	86	.471	
	1975	82	.545	

* Proportion of variation explained by the regression.

The implicit assumption is that the production of nursing homes should be measured in patient days rather than in cases treated. Since nursing homes generally provide ongoing supportive and residential services to individual patients with chronic disabling conditions, it would be inappropriate to use the episode or case as the unit of output.

would have been inaccurate and inappropriate to include the costs of capital. From the point of view of economic theory, accounting costs of capital are unlikely to reflect the opportunity cost of capital. In addition, reported costs reflecting expenditures for plant and equipment (depreciation, interest, rent) are unlikely to vary systematically with the independent variables reflecting scale and product mix, because these expense figures are determined by historic construction costs, age of the facility, method of financing and ownership types, as well as by the type and number of patient days provided. Moreover, there appears to be little substitution between capital and labor in the production of nursing home days, and reimbursement systems typically use different rules for reimbursement of capital and operating expenses.[3–5]

Summary of Results

Scale of Production

While average operating costs tend to decline as occupancy rate increases, i.e., as a facility is used to capacity, the studies find that the facility size, measured in number of beds or average daily census, is generally unimportant.¶ These results are summarized in Table 2.

The estimating procedure allowed costs to rise or fall over various size ranges by using Fuller's[5] "grafting" procedure. While average costs were found to fall initially as the bed size increases to about 40 beds, costs were not related to size beyond that level. Since diseconomies of scale were not present for intermediate to large nursing homes, these results do not support the notion that there is some optimal nursing home size. Relatively small facilities appear to have reached an efficient size.

The instability of the results over time for New York and Massachusetts suggests that the impact of regulation on facilities may be as important as underlying economies of diseconomies of scale and utilization. Over the period studied, rate-setting authorities varied the size groupings used to set rate limits in Massachusetts and New York, and differing penalties were applied to low-occupancy providers. These regulations apparently encouraged changes in provider behavior, as shown in Table 2. For the national analysis, which could not capture variation in state rate-setting treatment of size classes, size and utilization appear unimportant except for very small-scale effects over the lowest census range.

Provider Type: Ownership

The cost studies consistently found that facilities operated by nonprofit voluntary and government organizations have higher costs than for-profit nursing homes, by at least several dollars per day (Table 3). The studies have controlled for patient mix and services offered, and the New York study includes inspection-rating data that may be thought to control for quality of care. Nonetheless, the cost differentials between for-profit, nonprofit, and government providers persist. It is possible that the profit incentive encourages for-profit operators to be more efficient, so that they produce similar outputs with fewer inputs and lower costs.

Rate regulation appears to have played a role in producing these cost differentials, particularly in New York. The New York results for ownership differed over the three years studied. The New York rate-setting system grouped facilities by ownership in 1974 and 1975, so that nonprofit facilities knew their rates would be less limited than those of their for-profit counterparts. A nonprofit facility had little incentive to cut costs so long as costs were typical of nonprofit providers. For-profit and nonprofit facilities were grouped to-

---

¶ Average daily census is a measure of the rate of production equivalent to size (measured in number of beds) and capacity utilization (represented by occupancy rate) used in combination: average daily census equals occupancy rate times number of beds.

TABLE 2. Effect of Scale of Production on Average Operating Cost

	Independent Variables		Average Daily Consensus
	Beds	Occupancy Rate	
Sample	Effect per additional beds	Effect per additional % point	Effect per additional patient
National 1972	—	Insignificant	Negative then insignificant: costs fall by $0.20 per unit ADC for range 1–20 flat thereafter
Massachusetts 1973	Range 40–79: −$0.02	Range: 0– 90%: −$0.23 95– 97%: −$1.19 97–100%: +$0.46 Costs higher by $2.92 per day when occupancy is above 90%	
1974	Negative insignificant	Range: 0–90%: −$.0.21 90–93%: $1.23 95–97%: −$0.59 Costs lower by $1.69 per day when occupancy is above 93%	
1975	Negative insignificant	Range: 0–90%: −$0.12 95–97%: −$0.85 Costs higher by $0.161 per day when occupancy is above 95%	
New York 1974	Insignificant	Range: 80–90%: −$0.20 Costs higher by $4.89 when beds greater than 300	
1975	Insignificant	Insignificant	
1976	Range 1–49: $0.10 Costs less by $2.23 when 200–299 beds	Insignificant	
Indiana 1973	—	Insignificant	
1974	—	Insignificant	
1975	—	Insignificant	
National, 2 SLS 1972	Positive: +$0.10 per bed	Negative: −$0.02 per percentage point	

gether in 1976 for the first time, and the cost differential fell, decreasing from $5.96 to $4.14. This decrease in the otherwise increasing nonprofit/for-profit cost differential indicates that the nonprofit facilities have responded to the change in rate setting.

The nonprofit/for-profit differentials in Massachusetts between facilities were lower than those in New York. In this regard, it is significant that Massachusetts did not treat facilities differently by ownership for purposes of reimbursement.

Input Prices

The level of input prices clearly affects operating costs. The index of regional wage levels, proxied by the average retail wage in the county of facility location, was positively and significantly associated with the level of operating costs in the national and New York analyses. The county wage variable was not statistically significant in the Massachusetts analyses; however, counties in Massachusetts are much less compact than counties in other parts of the country and do not meaningfully represent labor market areas.

Another potential measure of input price variation was the geographic location of the home. In all cases, region was significantly associated with operating costs, as shown in Table 4. For example, in the national analyses, the per diem operating costs of homes in the Northeast were almost $4.00 greater than those of homes in the West; homes in the Central and South regions had costs of approximately $2.00 more than those of homes in the West. Analogous results were found for homes in different areas of Massachusetts and New York. It was expected that costs would be higher in urban areas, even after wage levels were accounted for, but this hypothesis was not supported in the single-equation national mode. However, in the two-stage least squares (2SLS) national model, which did not include regional location and found wage rates insignificant, SMSA location did have a significant effect on average operating cost. A further measure of input price, the actual average wage for licensed practical nurses in the facility, was included in the national and New York analyses. In both cases, this measure was positively and significantly related to operating costs.

TABLE 3. Effect of Ownership on Average Operating Cost Estimated Incremental Cost Relative to For-Profit Ownership

Sample	Nonprofit Voluntary	Government
National		
1972	$1.74	$2.52
Massachusetts		
1973	3.58	—
1974	3.66	—
1975	2.73	—
New York		
1974	7.08	9.43
1975	5.96	11.60
1976	4.14	10.30
National, 2SLS		
1972	+$1.36 for nonprofit and government combined	

Product Characteristics

The most difficult problem for nursing home cost studies is a proper description of the product of the nursing home. Cost function estimation can proceed only if the product is assumed to be homogeneous, or if product differences can be accounted for by including product mix descriptors among the independent variables. Three types of product descriptors have been used in these nursing home cost studies: 1) certified level of care; 2) measures of service availability and intensity; and 3) patient characteristics. Patient turnover and source of payment might also be treated as patient characteristics or product descriptors. These variables are discussed below.

Certified Level of Care. Certification status is a direct measure of product type, since skilled nursing facilities (SNFs) and intermediate care facilities (ICFs) are required to meet different standards. However, as discussed earlier, the definition and enforcement of these standards, while subject to federal guidelines, differ across states. Consequently, the services provided by SNF in one state may not be the same as those in another. Certification

TABLE 4. Effect of Location and Input Price Variables on Average Operating Cost

	Independent Variables		
Sample	Region	Urban Area: SMSA or Population Density	Wage Rates
National 1972	Northeast: + $3.62 Central: + $1.78 South: + $2.60	Insignificant	Retail wage, county: + $0.97 Nurse wage, facility: + $2.78
National, 2SLS 1972	—	$1.06	Insignificant
Massachusetts 1973	Significant cost differences among intrastate regions	—	—
1974			
1975			
New York 1974	Significant cost differences among intrastate regions	Insignificant	Retail wage, county: + $6.51
1975	"	Insignificant	Retail wage, county: + $6.29 LPN wage, facility: + $0.70
1976	"	Insignificant	Retail wage, county: + $6.83
Indiana 1973	—	Insignificant	Nursing home wage, county: + $3.59
1974	—	Insignificant	Insignificant
1975	—	Insignificant	Insignificant

level also proxies the type of patients, to the extent that patients are placed systematically according to need. Again, however, the effectiveness of the process that matches level of need to level of care apparently varies across states. The cost studies show that more complex (SNF) levels of care are indeed associated with higher costs (Table 5), but the differential varies across states.

Another reason for expecting differences between the costs of SNF and ICF providers is that, as with the size groupings and the nonprofit/for-profit distinction discussed above, certification status (SNF versus ICF versus combined SNF–ICF) is used by states as a grouping criterion for rate setting. In fact, this is the only grouping criterion required under federal guidelines for state rate setting.** As discussed above, when a criterion is used to group facilities, it is probable that cost differences between groups will be reinforced and perpetuated by the grouping itself. Given an initial difference in average cost between the groups, a rate-setting method that relies on group means sets rates or rate ceilings that differ between the groups by an amount approximating the initial average cost differences. There is no reason to expect this differential to diminish over time, since on average it will be paid by the rate-setting systems.

Further insight into the value of certified level of care as a predictor of cost is found

---

** During the period in question federal regulations required that SNF rates be higher than ICF rates.

in the 2SLS national model. Once nursing hours per patient day are accounted for, the SNF–ICF distinction does not add significantly to the explanation of cost. This implies that it is not level of care per se, but rather the associated nursing intensity that contributes to cost.

Output Descriptors: Services. Although information on services received by patients and services available was not complete, the service variables used did contribute to the explanatory power of the cost equations. Residual F-tests performed on the two data sets with relatively complete service information—the New York data and the national data—show that these variables are significant as a group, as shown in Table 6.

The endogenous service intensity variables included in the national 2SLS model, nursing hours per patient day and an index of rehabilitative-type services provided, were highly significant predictors of average operating cost. This result further supports the hypothesis that care differing in intensity differs in cost.

The inclusion of service intensity variables possibly raises methodological problems. While it is reasonable to assume that a day of nursing home care that includes three hours of licensed nursing care is a different product from a day with minimal nursing input, inclusion of a variable representing nursing hours per patient day could merely underline the obvious: nurses cost money (i.e., a day of care with higher nursing intensity costs more). Inclusion of variables for "intermediate" products, such as units of services provided to patients (for example, the per cent of patients receiving oxygen therapy, nasal feeding, catheterization, etc.) as in the national study, is perhaps theoretically preferable to direct inclusion of input variables like nursing hours. Equations including such service variables will estimate the average cost of providing the services as "add-ons" to the basic nursing home day of care. On the other hand, regulatory standards are often set up in terms of the minimally required nursing intensity, and the inclusion of that variable thus has practical importance.

TABLE 5. Effect of Certification as a Skilled Nursing Facility on Average Operating Cost

Sample	Differential Cost of SNF Over ICF Certification
National	
1972	$2.13
National, 2SLS	
1972	Insignificant
Massachusetts	
1973	$4.24
1974	$4.53
1975	$5.77
New York	
1974	$8.97
1975	$8.82
1976	$11.20
Indiana	
1973	$2.11
1974	$2.50
1975	$1.15

The results with respect to the number of beds per room, a proxy measure for quality, are relatively consistent across states. It costs more to provide care to patients in private rooms than to patients in rooms with two or more beds.

These results have clear implications for rate setting, for there are problems with using service-intensity variables as product descriptors. If more nursing hours change the product of a nursing home, a facility that uses nurses wastefully will not be identified as inefficient, but rather as producing an above average type of care. A facility providing excessive amounts of services, and perhaps keeping patients overly dependent, would be seen as providing a different product from more efficient counterparts that serve similar patients more effectively, but with fewer "doses" of service. If it could be assumed

TABLE 6. Effect of Services on Average Operating Cost

Sample	Independent Variables	
	Service Availability in Facility	Service Delivered to Patients
National		
1972	7 offered services significant as a group; for example: Occupational therapy (OT): + \$0.86 Physical therapy (PT): + \$0.91 Proportion of beds in rooms of 3 sizes: insignificant	13 variables representing % receiving specific services: insignificant as a group
Massachusetts		
1973	Insignificant	—
1974	PT + \$0.86 Care significantly less expensive in rooms with 4+ beds	—
1975	PT: + \$0.85 Care significantly more expensive in single rooms	—
New York		
1974	—	—
1975	Seven offered services: insignificant Care significantly more expensive in single rooms	8 variables representing number of services by type per patient day: significant as a group; e.g., +\$3.14 per PT visit per patient day
1976	—	8 variables representing number of services by type per patient day: significant as a group; e.g., \$1.25 per patient day
National, 2SLS	Proportion beds in rooms of three sizes: significant	Index of services delivered: significant Nursing hours per patient day: +\$3.29 per hour

that patients receive appropriate care accross facilities, patient status variables could be used to account for product differences, and there would be no need to include actual services provided. However, this has not been possible because of inadequate quality and patient status data.

Patient Characteristics. Conceptually, days of care provided to different types of patients are different products. Therefore, the cost studies used patient descriptors as product characteristics—so that the estimated cost functions predict the cost of basic nursing home care with cost "add-

TABLE 7. Effect of Patient Characteristics on Average Operating Cost

	Independent Variables				
Sample	Activities of Daily Living	Diagnoses	Mental States	Age	Sex
National					
1972	Insignificant	Significant	Depressed patients add to cost	Insignificant	Insignificant
National, 2SLS					
1972	Insignificant	Significant	—	Insignificant	Insignificant
Massachusetts					
1973	Insignificant	Insignificant	Insignificant	90+: $7.07	Insignificant
1974	Nonambulatory: +$1.73	Insignificant	Insignificant	90+: $3.26	Insignificant
1975	Nonambulatory: +$2.13	Insignificant	Insignificant	Insignificant	Insignificant
New York					
1974	Nonambulatory: +$3.34	—	—	Insignificant	Insignificant
1975	Eight variables significant as a group	—	—	Insignificant	—
1976	Insignificant	—	—	Insignificant	Insignificant

ons" for specific characteristics of the facility's patient mix.

The available patient descriptors were of four types: 1) age; 2) diagnosis; 3) level of dependence in activities of daily living (ADL); and 4) mental impairment. Patient differences may also be described by sex, but there was no a priori reason to expect cost differences between otherwise similar women and men. Increases in the magnitude of the case-mix variables can be hypothesized, in general, to correspond to increased dependence and need for care, and thus to increased cost. Table 7 summarizes the results.††

With one exception,‡‡ age variables, which might be expected to reflect debility that increases with age, were not found to add significantly to explanatory power. The proportion of female to male patients was never found to have a significant influence on average operating costs.

Measures of dependence in ADL, when considered as a group, did not add significantly to explanatory power in the national equation. However, econometric problems preclude strong inference from this result.

The New York measures of ADL appeared less complete, but did add significantly to explanatory power. When the number of patients requiring assistance with eating rises by one percentage point, average cost increases 4.5¢ per patient day; a one-percentage-point increase in those requiring assistance with bathing raised cost 3.2¢ per day; and a need for assistance with correspondence added 2.3¢ per day for each percentage point in patient mix.

In Massachusetts for 1974 and 1975, only one ADL-related variable was available, percentage of patients who were nonambulatory. Increasing the proportion of nonambulatory patients by one percentage point was found to increase cost by 1.5¢–3.5¢ per day.

---

†† The national data base had the largest array of patient descriptors. However, because the patient information there was based on a sample, each variable was subject to considerable error. The error would tend to bias the regression coefficients in the national equations toward zero. The absence of accurate quality measures also adds bias to the patient mix coefficient estimates.

‡‡ In Massachusetts in 1973 and 1974, it was found to be significantly more costly to care for those aged 90 or above.

Variables representing patient mix by diagnosis added significantly to the explanatory power of the estimated equation in the national analysis, but not in Massachusetts, where only a limited amount of diagnostic information was available. New York had no information on case mix diagnosis.

In summary, only a few kinds of patient status variables were significant in explaining cost, and the effect of particular variables is not consistent, either across years or across data sets. As in the case of services, this may be due to problems of measurement. However, the finding of general insignificance is also consistent with the hypothesis that resources expended for apparently similar patients differ widely across facilities. Cost per patient day depends more on the facility's ongoing staffing level and target level of quality than on patient needs determined in some objective sense. However, in the only data base that had measures of quality—New York—ADL variables are significantly associated with cost and consistent with the hypothesis that for the same level of "quality," a more debilitated patient is more costly.

Patient Turnover. Average length of stay may be seen as a patient variable, distinguishing short-stay patients recuperating from acute illnesses from patients who need long-term nursing and personal care. However, the variable may also describe the type of care being provided (maintenance versus rehabilitation) or the discharge policies of the facility.§§ Our studies show that high patient turnover is associated with higher per diem costs, either because of fixed costs per admission (administrative costs or the high cost of early days of care) or because of differences in patient or care type (Table 8).

Source of Payment. Conceptually, the source of payment for a given patient is a product descriptor only if patient needs or care provided vary systematically by payment source, and if the systematic differences have not already been controlled in the analysis. In any event, a facility's product type may vary by whether it serves mainly private or public (Medicaid) patients, or a mixture of these. Facilities oriented toward the private market may produce a different type of care, perhaps with a higher amenity level and commensurately higher costs than facilities serving primarily public patients. Alternatively, if private patients are more sensitive to price than public patients, and if rate setting for public patients is done on a full-cost reimbursement or pass-through basis, facilities serving more public patients may have higher costs. In addition, the per cent private pay available may reflect the effects of rate setting, utilization review and standards enforcement in facilities serving public patients.

Findings about the effect of the proportion of public-pay patients differ across the studies. (Table 9). In Massachusetts per diem costs were lower in public-pay-oriented homes, and in New York costs were higher. The coefficient of the payment proportion variable is insignificant in the national study. This result is not surprising, since the national results could average out the differences found in New York and Massachusetts. When an endogenous variable indicating per cent private pay was included in the national 2SLS model, with service intensity variables held constant, a strong negative effect on cost per patient day was observed. This result may be interpreted as showing that providers produce a given type of nursing home care (as indicated by the intensity

---

§§ Alchian[7] points out that with the volume of output held constant, average cost should fall as the number of production periods rises. In this context, length of stay may be interpreted as a proxy for the number of production periods.

TABLE 8. Effect of Admission Rate on Average Operating Cost

Sample	Estimated Coefficient* for Admissions Per Patient Day	Effect on Average Cost of Additional Admission for Typical Provider
National		
1972	$419	$0.017
National, 2SLS		
1972	$185	$0.007
Massachusetts		
1973	$377	$0.015
1974	$262	$0.010
1975	$113	$0.005
New York		
1974	$342	$0.014
1975	$1,050	$0.042
1976	$276	$0.011

* The coefficient on the variable "admissions per patient day" shows the effect on average cost of increasing the rate of admissions (that is, decreasing the average length of stay). The coefficients can be put into perspective in two ways. First, consider the effect on costs per patient day of an additional admission, which depends on the number of patient days. For the average nursing home surveyed by the National Center for Health Statistics in 1976, providing 23,962 patient days in that year, one more admission would have raised the admission rate from 0.00302 to 0.00306 (computed from NCHS, 1979). The effect of such a change on average cost per patient day for the cost studies is shown in the table. Alternatively, consider the effect on total annual costs of one more admission per year, again for the same number of patient days. The estimated equation shows the following relationship:

$$\text{average cost} = \frac{\text{total annual cost}}{\text{patient days per year}} = b \times \frac{\text{annual admissions}}{\text{patient days}} + \begin{matrix}\text{other}\\ \text{factors}\end{matrix}$$

where b is the coefficient of the admission rate. Multiplying through by patient days,

$$\text{total annual cost} = b \times \text{annual admissions} + \text{other factors}.$$

This implies that one more admission adds \$b to total annual costs. These values are shown in the estimated coefficient column.

variables) in a more efficient manner for private clientele who pay out of pocket than for public patients, for whom the government pays.

Interstate Differences in Reimbursement and Regulation

The national cost equations included variables describing differences in state regulation and reimbursement practices. In the single-equation model, nursing homes in states with prospective reimbursement were estimated to have higher costs, by $1.40 per patient day, than homes under cost-related or flat-fee reimbursement systems. The 2SLS national model, which accounts for simultaneous relationships among endogenous variables, also found that facilities in prospective reimbursement states had higher costs than those in cost-related reimbursement states. The coefficient for flat-rate reimbursement systems was insignificant, implying that flat-rate and prospective systems act similarly on costs. This result could be spurious for a variety of reasons. The result may be due to problems of equation specification and omitted variables—for example, omission of adequate quality measures. Alternatively, higher costs may initially be allowed by prospective systems to obtain nursing home operator approval of a system that places them at risk for unforeseen

TABLE 9. Effect of Private-Pay Patient on Average Operating Cost

Sample	Increment to Cost Per Percentage Point Private Pay*
National	
1973	Insignificant
National, 2SLS	
1972	−$0.020
Massachusetts	
1973	$0.025
1974	$0.045
1975	$0.051
New York	
1974	−$0.040
1975	−$0.036
1976	−$0.061
Indiana	
1973	Insignificant
1974	−$0.017
1975	Insignificant

* It would be misleading to interpret these results as showing the incremental cost of a private patient over that of a public patient in one facility. It may be assumed that public and private patients do not receive different care once they have been admitted to a facility; this may not be true if homes have separate public and private wings, but such practices are directly contrary to public policy and are not allowed under certification regulations. Therefore, the coefficient of percentage private pay is believed to proxy the commitment to the private market.

events. Or, it may be that the cost ceilings used in retrospective systems are more cost constraining than the rates set under prospective reimbursement. Another explanation might be that those states with excessively high nursing home costs adopted the prospective approach in an effort to "solve" their cost problem. An appropriate test for the effect of prospective reimbursement would be to observe homes before and after statewide adoption of prospective reimbursement.

The single-equation cost model also found that other state regulations appear to affect the costs of nursing homes. Homes in states with life-safety codes have higher costs, by about $0.75; nursing homes in states with inflexible nursing staff requirements have higher costs, by $1.45; similarly, occupancy penalties significantly and positively affect costs. Several other regulatory variables—including the presence of a certificate of need program, preadmission patient assessments and a uniform chart of accounts—have positive but statistically insignificant cost impacts. Finally, reimbursement systems that put limits on total payments to nursing homes or on administrative salaries may be successful in reducing costs. However, the negative coefficient was not statistically significant.

Some of these policy variables achieved significance in the 2SLS cost equation, which did not include region variables. For example, certificate-of-need programs were found to be significantly related to higher costs (by $2.43 per day), as were the staffing requirement variables and the requirement of a uniform chart of accounts. These findings suggest that regulation and reimbursement fall into regional patterns and are probably related to and caused by the previous cost, access and quality performance of the state's nursing home industry. This suggests that analyses of the association between cost and regulatory practice at one point in time may be unreliable for predicting the likely effect elsewhere.

In concluding this discussion, some overall comments on the results of the multi-equation 2SLS results are appropriate. The results underline the extent to which simultaneity issues pervade analysis of nursing home behavior. It is found, for example, that an increase in service intensity increases operating costs, but also that it increases the occupancy rate and the percentage of private patients in a home, and that both of these latter effects are cost-decreasing. Furthermore, regulation and rate setting are seen to have a variety of effects on policy-relevant variables other than cost. In particular, reimbursement variables also affect service intensity, to rates charged to private patients and the public–private mix of patients. The demand for nursing home care by private-pay patients was found to be elastic with respect to price. In addition, private-pay patients were found to be paying higher rates than public patients, and cost-controlling reimbursement features were found to increase the differential. In other words, private patients appear to be subsidizing public patients. The conclusion, then, is that appropriate rate-setting practices cannot be formulated directly on the basis of cost analyses. Rather, the results underline the importance of a more comprehensive understanding of the nursing home environment if unintended and undesirable consequences of rate setting are to be avoided.

References

1. Bishop CE. Nursing home cost studies and reimbursement issues. Health Care Financing Review 1980;1(4):47.

2. Birnbaum, H, Lee AJ, Bishop C, Jensen G. Public pricing of nursing home care. Cambridge, MA: Abt Books, 1981.

3. Battelle Memorial Institute. Profit and growth: allowances for owner/operator contributions in a cost related prospective payment for services system for nursing home care. Washington, D.C.: American Health Care Association, 1975.

4. Schulman D, Galanter R. Reorganizing the nursing home industry: a proposal. Milbank Mem Fund Q 1976;54:129.

5. Winn S, Bennett C, McCaffree KM. The relationship of resident patient condition and characteristics to variations in costs of long term care. (Mimeographed.) Seattle, WA: Battelle Health Care Study Center, 1976.

6. Fuller A. Grafted polynomials as approximating functions. Australian Journal of Agricultural Economics, June 1969.

7. Alchian A. The allocation of economic resources. Palo Alto, CA: Stanford University Press, 1959.

Medical Care

December 1981
Vol. 19 No. 12
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Out-of-Plan Use Under Two Prepaid Plans

Membership Duration and Utilization Rates in a Prepaid Group Practice

Determining the Generalizability of Rating Scales in Clinical Settings

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
The Rand Corporation

Gordon H. De Friese, Ph.D.
University of North Carolina

Brant E. Fries, Ph.D.
Yale University

Paul M. Gertman, M.D.
Boston University Medical Center

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Joel W. Kovner, Dr.P.H.
Kaiser Permanente Medical Care Programs, Los Angeles

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

Steven A. Schroeder, M.D.
University of California, San Francisco

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $40.00; special price U.S. students, $23.00; all other countries except Japan, one year, $50.00. Single copies $5.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1981 by J. B. Lippincott Company. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

December 1981
Volume 19, Number 12

CONTENTS

EDITORIAL

On the Importance of Exploring the Data. *Edward B. Perrin* 1163

ORIGINAL ARTICLES

Out-of-Plan Use Under Two Prepaid Plans. *Anne A. Scitovsky, Lee Benham and Nelda McCall* 1165

Membership Duration and Utilization Rates in a Prepaid Group Practice. *Mary Jane Griffith and Nicholas Baloff* 1194

Determining the Generalizability of Rating Scales in Clinical Settings. *William J. Evans, C. Gene Cayten and Paul A. Green* 1211

COMMUNICATIONS

Commentary on "Membership Duration and Utilization Rates in a Prepaid Group Practice," by Mary Jane Griffith and Nicholas Baloff. *John P. Mullooly* 1221

Commentary on "Membership Duration and Utilization Rates in a Prepaid Group Practice," by Mary Jane Griffith and Nicholas Baloff. *Charles E. Yesalis, III and Susan D. Horn* 1226

Reply to Comments: "Membership Duration and Utilization Rates in a Prepaid Group Practice." *Nicholas Baloff and Mary Jane Griffith* 1233

An Organized Locum Tenens Service and the Cost of Free Time for Rural Physicians. *Alan K. Kronhaus* 1239

BOOK REVIEWS

Unequal Care: A Case Study of Interorganizational Relations in Health Care, by Murray Milner, Jr. *Gordon H. DeFriese* 1247

Working for a Healthier America, edited by Walter J. McNerney. *David Cohen and Duncan Neuhauser* 1248

INDEX TO VOLUME 19 1250

MEDICAL CARE
December 1981, Vol. XIX, No. 12

Original Articles

Out-of-Plan Use Under Two Prepaid Plans

ANNE A. SCITOVSKY, M.A.,* LEE BENHAM, PH.D.,† AND NELDA MCCALL, M.A.‡

This analysis of out-of-plan use of physician and paramedical services under a Kaiser plan and under a prepaid option offered by a predominantly fee-for-service group practice (Clinic plan) deals primarily with services that members could have obtained from plan providers ("covered services"). The extent and pattern of out-of-plan use were found to be similar. While 16–20 per cent of plan members used some out-of-plan covered services and the mean number of such services was about one half visit per member per year, most out-of-plan users were occasional users, 10–12 per cent of users (or 2 per cent of plan members) accounting for 50 per cent of all out-of-plan covered services. The principal member characteristics associated with out-of-plan use were dissatisfaction, health status and having other insurance. The literature on out-of-plan use is also reviewed.

NO EVALUATION OF THE PERFORMANCE of a prepaid health plan, or Health Maintenance Organization (HMO), and especially no comparison of the performance of different plans, is complete without an exploration of out-of-plan use of medical services by plan members. If there is substantial out-of-plan use, or greater out-of-plan use under one plan than another, the in-plan utilization figures are misleading. Even more important, it has been suggested that out-of-plan use may be the reason that HMOs are able to provide medical care more economically than the fee-for-service system. Members of HMOs, so the argument goes, get their day-to-day medical care from the HMO but in case of more serious illness consult nonplan physicians.

In this article, we analyze out-of-plan use under two prepaid plans, on which we conducted a study to explore the performance of two different HMO-type organizations. Other findings of this study have been published in previous issues of this journal.[1–3] The principal questions that this article addresses are the magnitude of out-of-plan use in terms of the number of out-of-plan users as well as of out-of-plan services, the characteristics of out-of-plan users, the types of out-of-plan ser-

---

* Chief, Health Economics Division, Research Institute of the Palo Alto Medical Foundation, Palo Alto, California; and Lecturer, Health Policy Program School of Medicine, University of California, San Francisco, California.

† Department of Economics and Division of Health Care Research (Preventive Medicine), Washington University, St. Louis, Missouri.

‡ Manager, Health Policy Research, SRI International, Menlo Park, California.

The authors wish to express their gratitude to Mary Cline for her help with all tables and her careful checking of all figures.

This study was supported by Contract No. OA-OFM-SSA-PMB-75-238 from the Social Security Administration and the National Center for Health Services Research, and by Grant No. 18-P-97027/9-01 from the Health Care Financing Administration.

Address for reprints and correspondence about manuscript: Anne A. Scitovsky, Research Institute of the Palo Alto Medical Foundation, 860 Bryant Street, Palo Alto, CA 94301.

0025-7079/81/1200/1165/$01.95 © J. B. Lippincott Co.

vices used and out-of-pocket expenses for out-of-plan services. The article also discusses the problem of defining and measuring out-of-plan use, which has troubled the literature on the subject, and compares our findings with those of other studies of out-of-plan use.

Description of the Plans

Since the two plans have been described in detail in our previous articles,§ only a brief summary of their main features will be given here. They are basic health care plans offered by Stanford University to its employees and their families.‖ One is a Kaiser plan, the other is United Medical Clinics/Blue Cross plan (hereafter referred to as Clinic plan). Under this plan, physician services and outpatient ancillary services are provided by the Palo Alto Medical Clinic (PAMC), a multispecialty, predominantly fee-for-service group practice of about 125 physicians, and hospital services are covered by a Blue Cross policy. The benefit provisions of the two plans are almost identical, covering most medical services in and out of the hospital, the main exclusions being psychiatric care and, in the case of the Clinic plan, routine eye examinations. The main difference between the plans, in addition to the difference in the organizational structure of the providers, lies in their cost-sharing provisions. While the Kaiser plan has no cost sharing beyond a $1 charge for a physician office visit and $3.50 for a home visit, the Clinic plan has a 25 per cent co-insurance provision that applies to all physician services and to outpatient ancillary services. This means that the marginal cost of going outside the plan for a service obtainable under the provisions of the plan is lower for Clinic plan members than for Kaiser plan members. Other things being equal, we would therefore expect greater use of out-of-plan services by Clinic than by Kaiser plan members.

In addition to its basic health plans, Stanford University also provides a major medical insurance policy that employees obtain free of charge as a fringe benefit. Dependents can be enrolled for a monthly premium, which at the time of our study was $5 if they were members of one of the basic health plans and $7 if they were not members. After a deductible of $100 for basic plan members and $500 for nonmembers, it pays 80 per cent of most medical and hospital expenses, including some costs not covered by the basic plans, such as drugs and prescriptions. Enrollees in the basic health plans are required to use these plans for covered services before resorting to their major medical coverage. The policy will, however, pay for such services when obtained from an out-of-plan source, but only after a deductible of $300—rather than the $100 applicable to basic plan members for other medical expenses—has been met.

Study Population and Methodology

The study population consists of 2,061 Kaiser and 2,139 Clinic plan members who were covered by the plans during the full 12 months from July 1973 through June 1974 and who in addition participated in an initial household interview and three follow-up telephone interviews over the 12-month period. The selection of this population has been described in detail in one of our earlier articles.[1] The distribution of the two groups by various demographic, socioeconomic and health-related characteristics, also described at length in our earlier articles,[1,2] is shown in Table 1. The principal differences between the two groups that may be of note here are the higher socioeconomic status of Clinic plan members in terms of family income, education and occupation; the higher proportion of older members and of members covered by Stanford's major medical insurance among Clinic plan enrollees; the larger percentage of dissatisfied Kaiser plan members and of Kaiser plan members covered by additional insurance other than Stanford's major medical insurance; and the shorter average length of plan membership of Kaiser plan members. There is little difference, however, in the health status of the two groups as measured in terms of self-assessment, per capita number of in-plan physician (and physician plus paramedical) services, and number of hospital days per 1,000 plan members.#

---

§ See especially reference 1.

‖ Stanford University also offers a third basic plan, a standard Blue Cross medical/hospital policy, which is not included in this study.

Data were collected for the 12-month period from July 1973 through June 1974. While data on in-plan use were obtained from the plan members' medical records, data on out-of-plan use were collected in an initial household interview and three follow-up telephone interviews held at 3-month intervals. Respondents who reported an out-of-plan service were asked the date of the service, the name of the provider (so that it was possible to establish whether the provider was a physician or some other health service professional), the reason for going outside the plan, the condition for which out-of-plan care was sought and the type of service obtained, the place of the service, the number of visits and out-of-pocket expenses, if any. The initial household interviews also furnished information on marital status of the subscriber, family income, education and occupation of the head of the household, additional insurance coverage, attitudes concerning medical care, satisfaction with the plans, health status and regular source of care.

Definition and Measurement of Out-of-Plan Use

Before discussing our findings, it is necessary to say a few words about the problem of defining and measuring out-of-plan use. While there have been quite a few studies of out-of-plan use under prepaid plans,[4-22] either dealing specifically with the subject or touching on it within the framework of a broader discussion of particular plans, the results have on the whole been disappointing and confusing.¶ The reason for this unsatisfactory state of affairs is that there exists neither a common definition of what constitutes out-of-plan use nor agreement on how to measure it.

In many studies of out-of-plan use, it is not made clear what the term encompasses. In most of them it seems to be implied that an out-of-plan service is any service obtained by a plan member from a nonplan physician and, in some instances, from some other type of health professional, such as a chiropractor. A few studies actually define it as such.[11,19] Thus, in most studies, three different types of out-of-plan use are combined: 1) services not covered by the plans and not paid for by them; 2) services covered by the plans that plan members obtained from nonplan providers but that the plans paid for (e.g., services when members were temporarily out of the area, emergency services, and referrals by the plans to outside providers); and 3) services covered by the plans that plan members elected to obtain from nonplan providers and for which the plans did not pay. Adding to the confusion is the fact that some studies appear to be limited to out-of-plan use of physician services, while others include the services of other health professionals, and that telephone consultations are included in some studies but excluded in others.

For our study of hospital utilization, we used a larger study population; see reference 3.

¶ For a good review of the literature, see reference 17.

TABLE 1. Selected Characteristics of Kaiser and Clinic Plan Members

Member Characteristics	Percentage of All Members[*]			
	Kaiser	(N)	Clinic	(N)
All members	100	(2061)	100	(2139)
Sex				
Male	49	(1014)	48	(1036)
Female	51	(1047)	52	(1103)
Membership status				
Subscriber				
Male	29	(590)	29	(625)
Female	16	(336)	12	(265)
Spouse				
Male	2	(50)	2	(34)
Female	17	(344)	21	(453)
Dependent				
Male	18	(374)	18	(377)
Female	18	(367)	18	(385)
Age				
Under 15	28	(581)	27	(587)
15–44	53	(1094)	41	(882)
45–64	18	(377)	31	(663)
Unknown	<1	(9)	<1	(7)
Marital status of subscribers and spouses[†]				
Married subscriber	31	(643)	32	(674)
Married spouse	19	(388)	23	(483)
Never married	7	(142)	5	(97)
Widowed, divorced, separated	7	(147)	6	(123)
Family income				
Under $10,000	22	(452)	8	(157)
$10,000–24,999	71	(1470)	53	(1130)
$25,000+	6	(123)	40	(833)
Unknown	<1	(16)	<1	(19)
Occupation of household head				
Faculty	8	(167)	39	(843)
Other professional staff	40	(832)	45	(960)
Nonprofessional staff	52	(1061)	16	(336)
Unknown	<1	(1)	—	(—)
Education of household head				
High school graduate or less	23	(467)	7	(140)
Some college or college graduate	41	(846)	23	(491)
Some postgraduate work or graduate degree	36	(744)	70	(1507)
Unknown	<1	(4)	<1	(1)
Covered by Stanford Major Medical Insurance				
Yes	66	(1352)	93	(1990)
No	34	(709)	7	(149)
Covered by outside insurance other than Stanford Major Medical				
Yes	12	(237)	9	(189)
No	88	(1824)	91	(1950)

TABLE 1. (Continued).

Member Characteristics	Percentage of All Members*			
	Kaiser	(N)	Clinic	(N)
Satisfaction with the plan				
Completely satisfied	28	(571)	36	(770)
Generally satisfied	61	(1249)	58	(1248)
Not satisfied	11	(223)	6	(118)
Unknown	<1	(18)	<1	(3)
Health status				
Excellent	59	(1209)	63	(1351)
Good	34	(709)	31	(665)
Fair	6	(130)	5	(106)
Poor	<1	(11)	<1	(14)
Unknown	<1	(2)	<1	(3)
Had one or more hospitalizations in 12-month period				
Yes	3	(66)	4	(82)
No	97	(1995)	96	(2057)
Number of in-plan physician and paramedical visits for illness in 12-month period				
None	31	(638)	28	(590)
1–2	32	(651)	32	(690)
3–5	20	(419)	23	(501)
6 or more	17	(353)	17	(358)
Number of family members covered				
1	23	(466)	16	(342)
2	15	(308)	18	(384)
3–4	37	(771)	45	(961)
5 or more	25	(516)	21	(452)
Number of years in plan				
1	7	(140)	3	(59)
2–4	46	(950)	21	(444)
5–9	47	(971)	25	(545)
10 or more	—	(—)	51	(1091)
Regular source of care				
Specific plan physician	42	(870)	87	(1865)
Plan, no specific physician	51	(1051)	10	(211)
Outside physician	3	(52)	1	(23)
No regular source of care	4	(88)	2	(40)

---

* Per cents may not add to 100 because of rounding.

† Dependent children are not included.

One of the few studies in which it is clearly stated which type of out-of-plan use is being examined is the study by Pope et al.[18] dealing with outside use of physicians by members of the Portland, Oregon, Kaiser plan. The authors distinguish four types of out-of-plan services: 1) services neither covered by the plan nor available at Kaiser; 2) services not covered but available at Kaiser; 3) services covered by the plan but not available at Kaiser, for which physicians outside the Kaiser system are utilized; and 4) services available at Kaiser and covered if received there. It is this last category—services for which a "decision was made *not* to use Kaiser though the services were available and would have been covered"—with which their study deals. The authors add, "This category would logically include those services which have been received, for example, at work or school as a matter of company or school policy." They do not include these in their study, however, mainly because only about 1 per cent of the study population reported some service of this type during the 12 months covered by the study.

In addition to the problem of how to define out-of-plan use, there is the problem of how to measure it. Luft[17] shows this problem to have three major aspects: the unit of measurement, the unit of observation and the time period covered. Some studies use outside users as the unit of measurement, others outside services, with only a few looking at both. Similarly, some studies use the family as the unit of observation, others the individual plan member. Finally, some studies examine outside use during a specific time period, such as the previous year, while others refer to outside use during the entire period of an individual's membership in the plan. It is not surprising that, as a result of these definitional and measurement problems, the picture of outside utilization under prepaid plans that has emerged from the existing studies is blurred and confusing.

In this study, we define out-of-plan use as all medical services other than telephone consultations obtained from nonplan physicians and paramedical personnel, except those for which the plans paid (i.e., emergency services, services when members were temporarily out of the area and referrals by the plans). We distinguish between two types: 1) out-of-plan services not covered by the plans and not paid for by them and 2) out-of-plan services that would have been covered by the plans had they been obtained from plan providers but that the plan member elected to obtain from an outside provider and for which the plans did not pay. The principal services in the first category are psychiatric services, some paramedical services not covered by the plans, notably chiropractors' services, and, in the case of the Clinic plan, routine eye examinations. We shall refer to the first type of out-of-plan services as noncovered services and to the second as covered services.

In the case of covered out-of-plan services, we make a further distinction between those services for which plan members paid and those for which they did not pay. For want of a better term, we shall refer to the second group as "free" services, while the term "covered" services will include both "free" and paid-for services. "Free" services in this study consist largely of services obtained at public clinics, free of charge because of professional courtesy, at no cost under public programs (e.g., Worker's Compensation, Veterans Administration and CHAMPUS), and by employees of the Stanford Linear Accelerator (SLAC) at its on-premise facility.** They also include some services paid

---

** Our study population of 4,200 plan members includes 323 SLAC employees, of whom 212 were Kaiser plan members (representing 10.3 per cent of the Kaiser plan population) and 111 were Clinic plan members (representing 5.2 per cent of the Clinic plan population). We duplicated the entire analysis excluding the SLAC employees and found that this did not alter our principal findings.

for by other insurance, such as duplicate policies obtained by plan members through their spouses' employment, student health policies of dependent children and a few instances of insurance of other persons (e.g., liability insurance of another person in cases of accidents).

Our analysis will focus primarily on out-of-plan use of covered physician and paramedical services. While out-of-plan use of noncovered services is of interest mainly because it indicates the extent to which the benefit package of a plan meets—or fails to meet—the needs of plan members, out-of-plan use of covered services raises questions about both the performance and the membership of prepaid plans. The basic question is why persons who join a comprehensive prepaid plan would go to nonplan providers at all for services they could obtain from plan providers at no cost, or at no cost beyond the copayment required by the plan. Is such out-of-plan use mainly an indication of dissatisfaction with the plan, or does it represent something else? What are the characteristics of persons who go outside the plan for services they could obtain from plan providers? Who are the heavy users of such out-of-plan services? What are the reasons given by out-of-plan users for going outside the plan? These are some of the questions we shall try to answer.

The analysis will also be carried out largely in terms of all covered physician and paramedical services, including "free" services. In contrast with the findings of Pope,[18] we found (as can be seen in Table 3) that the extent of out-of-plan use of "free" covered services was by no means negligible, either in terms of the percentage of members with such out-of-plan services or of the number of such services per member. Our principal interest is in the *total* use of medical services by members of the two plans. Omitting the "free" services would therefore have given an inaccurate picture of overall use, since it is quite likely that in their absence, in-plan use or other out-of-plan use would have been higher, though not necessarily by the full amount of the "free" services. However, we have duplicated the entire analysis for covered out-of-plan services excluding "free" services and present the results in our tables.

We have included services by paramedical personnel in our analysis because, as in the case of "free" services, their omission would have given a somewhat distorted picture. This is because Kaiser plan members obtained a substantially higher proportion of their out-of-plan services from paramedical personnel than did Clinic plan members. The relevant figures are shown in Table 2. It is clear that out-of-plan use of covered services by Kaiser plan members, as compared with that of Clinic plan members, would have been understated if paramedical services had not been included.

Our analysis will be in terms of both out-of-plan users and out-of-plan services in the 12-month period from July 1973 through June 1974. After an overview of our findings, we shall begin by examining the out-of-plan users: how many are there, what are their characteristics, what differences are there between the two groups? This discussion will be followed by an analysis of out-of-plan services: how many are there, how are they distributed among different users, who are the heavy users? Next will come a review of the reasons

TABLE 2.

Per cent of Covered Out-of-Plan Physician and Paramedical Services Obtained from Paramedical Personnel

	Kaiser Plan Members	Clinic Plan Members
All covered services	10.1	4.2
Covered excluding "free" services	6.3	.5
"Free" services	15.5	10.2

given by plan members for going outside the plan for covered services. The use of noncovered services will be dealt with briefly in the next section, followed by a short section on out-of-pocket expenses for both covered and noncovered services. In a final section, we shall evaluate our findings in the light of those of other studies on out-of-plan use.

Findings

Overview

The basic findings of the study are summarized in Table 3. It shows that 25.2 per cent of Kaiser plan members and 31.7 per cent of Clinic plan members had at least one out-of-plan covered or noncovered physician or paramedical service during the 12-month period of the study. If Clinic plan members whose only out-of-plan service was a routine eye refraction (which is a covered service under the Kaiser plan but a noncovered service under the Clinic plan) are omitted, the Clinic plan percentage drops to 20.8 per cent, slightly below the Kaiser percentage. The percentage of out-of-plan users of covered services was somewhat higher among Kaiser than Clinic plan members, with 20.3 per cent of Kaiser and only 15.6 per cent of Clinic plan members reporting at least one such service. This difference also holds true when we separate covered services into "free" services and services for which members paid a fee; 11.5 per cent of Kaiser and 9.1 per cent of Clinic plan members went outside the plan for at least one covered service for which they had to pay, with corresponding figures for "free" services of 9.9 per cent and 7.4 per cent, respectively. Thus, our hypothesis that Clinic plan members would be more likely to go outside the plan for covered services is not borne out by the data. While a substantially higher percentage of Clinic than Kaiser plan members used noncovered services—19.7 per cent compared with 6.7 per cent—the Clinic plan figure is 6.0 per cent if Clinic plan members whose only noncovered service was a routine eye refraction are excluded.

In terms of the number of out-of-plan physician and/or paramedical services per 100 plan members during the 12-month period, total out-of-plan use of covered and noncovered services by the two groups was

TABLE 3. Per cent of Kaiser and Clinic Plan Members With at Least One Out-of-Plan Physician or Paramedical Service and Number of Out-of-Plan Services (July 1973 to June 1974)

	% with at Least One Out-of-Plan Service		No. of Out-of-Plan Services per 100 Members	
Type of Service	Kaiser	Clinic	Kaiser	Clinic
	100.0 (N = 2061)	100.0 (N = 2139)		
All out-of-plan services	25.2	31.7 (20.8)*	163	165 (148)†
Covered services	20.3	15.8	62	50
Covered services excluding "free" services	11.5	9.1	36	31
"Free" services	9.9	7.4	26	19
Noncovered services	6.7	19.7 (6.0)*	101	115 (98)†

* Excludes members whose only noncovered out-of-plan service was a routine eye refraction.

† Excludes routine eye refractions.

almost identical, 163 such services for Kaiser and 165 for Clinic plan members. However, if routine eye refractions of Clinic plan members are excluded, the Clinic plan figure is reduced to 148 services per 100 members. As in the case of the percentage of members who used at least one out-of-plan covered service, Kaiser plan members were also somewhat heavier users than Clinic plan members in terms of the number of such services per 100 members, with the Kaiser figure of 62 compared with the Clinic figure of 50. Again, our hypothesis that Clinic plan members would be heavier out-of-plan users is not borne out by our findings. The difference between the two groups is somewhat more pronounced in the case of "free" services than of services for which members had to pay. Kaiser plan members had 26 "free" services per 100 plan members, as compared with only 19 for Clinic plan members, while the figures for out-of-plan covered services for which members had to pay were closer, amounting to 36 and 31 per 100 members, respectively. Finally, while Clinic plan members used more noncovered services (115 compared with 101 per 100 members), their rate of use of such services is much the same as the Kaiser rate when routine eye refractions are excluded from the Clinic figures.

When the out-of-plan use data are placed in the context of total medical care use of the two groups, as is done in Table 4, which combines our data on in-plan and out-of-plan use of physician and paramedical services, the similarities between the two groups are striking. The per capita number of all medical services was practically identical: 5.58 for Kaiser and 5.55 for Clinic plan members. Both groups obtained about 70 per cent of all their physician and paramedical services from plan providers and 30 per cent from nonplan providers. Just over 11 per cent of the services obtained by Kaiser plan members and 9 per cent of those obtained by Clinic plan members were services obtained from out-of-plan providers that could have been obtained from plan providers under the terms of the plans.

In addition to out-of-plan physician and paramedical services, both groups also had a small number of out-of-plan hospital stays. Kaiser plan members had nine (4.4 per 1000 members) and Clinic plan members eight (3.7 per 1000 members) hospital admissions for noncovered services; of these, five each were psychiatric hospitalizations. While the Clinic group had no out-of-plan hospital admissions for covered services, the Kaiser group had seven (3.4 per 1000 members), totalling 32 hospital

TABLE 4. Per Capita Number and Percentage Distribution of In-Plan and Out-of-Plan Physician and/or Paramedical Services (July 1973 to June 1974)*

Type of Service	Per Capita No. of Services				% of Services			
	Kaiser		Clinic		Kaiser		Clinic	
In-plan services	3.95	(3.59)	3.90	(3.83)	70.8	(76.5)	70.3	(74.4)
Out-of-plan services	1.63	(1.10)	1.65	(1.32)	29.2	(23.5)	29.7	(25.6)
Covered	.62	(.56)	.50	(.48)	11.1	(11.9)	9.0	(9.3)
Noncovered	1.01	(.54)	1.15	(.84)	18.1	(11.5)	20.7	(16.3)
TOTAL	5.58	(4.69)	5.55	(5.15)	100.0	(100.0)	100.0	(100.0)

* Figures in parentheses refer to physician services only.

days (15.5 per 1000 members). Reasons given for these hospitalizations were emergency care (two cases), "too long to get appointment" (two cases), "too far away" (one case) and "preferred outside physician" (two cases).

Covered Services: Characteristics of Users

To turn to the detailed analysis of out-of-plan use of covered services, Tables 5 and 6 present our findings on the characteristics of out-of-plan users of these services. Table 5 shows the percentage of plan members with at least one out-of-plan covered physician or paramedical service by selected member characteristics. The figures in the first two columns refer to plan members with either type of out-of-plan covered service, paid for or "free," and, as stated earlier, our analysis will focus on this group. However, we also show, in the other two columns, the same data excluding plan members whose only out-of-plan covered services were "free" services. Table 6 presents the results of the multivariate analysis.†† The dependent variable is a dummy variable that is equal to one if the individual had one or more out-of-plan physician or paramedical services and zero if he had no such services. The independent variables are the member characteristics shown in Table 5 except regular source of care, occupation and education.‡‡ All of these except number of family members covered and number of in-plan physician and paramedical services for illness are separated into discrete intervals and enter the equation as dummy variables. Within each set of dummy variables, one category is excluded (e.g., male subscriber), and the coefficients of the remaining dummy variables in each set can be regarded as indicating the probability of being an out-of-plan user with reference to the excluded category. As in Table 5, the first two columns refer to out-of-plan users of all covered services, the other two only to those members who used services for which they had to pay.

Because the cross tabulations of Table 5 and the multivariate analysis of Table 6 show much the same association between member characteristics and being an out-of-plan user, we shall discuss this subject mainly on the basis of the multivariate analysis. It should be noted, however, that Table 5 also shows not only that the Clinic plan group as a whole had a smaller proportion of out-of-plan users than the Kaiser plan group (which we already noted in Table 3) but that this was true for most of the subgroups as well. Among the few exceptions are plan members who reported their health as fair or poor, and members who had one or more hospitalizations in the 12-month period. Clinic plan members with these characteristics had a considerably higher percentage of out-of-plan users than the corresponding Kaiser plan groups.

The multivariate analysis presented in Table 6 shows that for both groups, holding other factors constant, the member characteristics significantly associated with going outside the plan for covered services were satisfaction with the plans, having outside insurance, number of family members covered and length of plan membership. In addition, several other characteristics—membership status, age, marital status and health status—showed some significant relationship with outside use for one group but not the other.

For both groups, satisfaction with the plan was one of the most important factors associated with being an out-of-plan user. Kaiser plan members who were not satisfied were 16.5 per cent more likely and Clinic plan members 10.8 per cent more likely to be out-of-plan users than members who were completely satisfied. These differences are not only statistically significant but relatively large, considering that the average Kaiser plan member had only a 20.3 per cent and the average Clinic plan member a 15.8 per cent probability of being an out-of-plan user. Even plan members who were generally (but not completely) satisfied were slightly but significantly more likely to use out-of-plan services, the probabilities being 5.7 per cent for Kaiser and 3.6 per cent for Clinic plan members.

---

†† We also calculated logit estimates in place of one of the equations and found that they gave very similar results.

‡‡ We excluded regular source of care because of colinearity between having an out-of-plan physician as regular source of care and using an out-of-plan physician. We omitted occupation and education because after running various regressions using different combinations of the three socioeconomic variables, we found that it made little difference which variable we used, and we selected family income.

Having additional health insurance other than Stanford University's major medical insurance is another factor strongly and significantly associated with going outside the plan for both groups. Kaiser plan members without such additional insurance were 11.1 per cent less likely and Clinic plan members 10.2 per cent less likely to be outside users than members who had such insurance. It is not clear why coverage by Stanford University's major medical insurance was found to have no relationship to out-of-plan use. One possible explanation is member ignorance. Stanford University employees who work at least 75 per cent of full time get this insurance automatically and at no cost to themselves. The household interviews revealed that a surprising number of subscribers were unaware that they were covered by it. In addition, the $300 deductible applicable under the major medical policy when plan members obtain a covered service from a nonplan provider makes the policy irrelevant for the occasional out-of-plan use:.

The data in Table 6 also show a small but significant inverse relationship between number of family members covered and going outside the plan. For every additional family member covered, the probability of being an out-of-plan user declined 2.0 per cent in the case of Kaiser plan members and 1.5 per cent in the case of Clinic plan members. A similarly small but significant relationship was also found between length of plan membership and out-of-plan use. Members of both plans who had been covered by them less than five years were 4.7 per cent more likely to use some outside services than those who had been members longer.

The relationship between membership status and being an out-of-plan user is somewhat different for the two groups. Both Kaiser and Clinic plan male spouses were the least likely to be out-of-plan users, the probability being 21.0 per cent less for Kaiser and 14.1 per cent less for Clinic plan male spouses than for male subscribers. But while Kaiser plan female spouses and male and female dependents also were significantly less likely to go outside the plan, this was not found to be the case for the corresponding Clinic plan groups. On the other hand, while Kaiser plan male and female subscribers were equally likely to be outside users, Clinic plan female subscribers were 7.8 per cent more likely to use outside services than male subscribers, a statistically significant difference.

The relationship between age and being an outside user is also slightly different for the two groups. While in both groups the youngest age group (under 15 years) was found to be less likely to go outside the plan than the oldest group (45–64 years), the difference is statistically significant only for Kaiser plan members. But in both groups, there was no difference in the probability of out-of-plan use between the 15–44 age group and the oldest group. Marital status and being an out-of-plan user showed no relationship for Clinic plan members, but Kaiser plan never-married members were 10.3 per cent more likely to go outside the plan than married members, a statistically significant difference, while the probability of out-of-plan use was the

TABLE 5. Per Cent of Kaiser and Clinic Plan Members with at Least One Out-of-Plan Covered Physician or Paramedical Service (July 1973 to June 1974)

Member Characteristics	Covered Services		Covered Excluding "Free" Services	
	Kaiser	Clinic	Kaiser	Clinic
All members	20.3	15.8	11.5	9.1
Sex				
Male	20.3	15.6	8.9	7.4
Female	20.3	16.0	14.0	10.7
Membership status				
Subscriber				
Male	27.6	17.4	9.5	7.8
Female	33.3	14.7	23.8	8.3
Spouse				
Male	8.0	8.8	4.0	8.8
Female	18.0	19.0	12.8	15.0
Dependent				
Male	10.4	13.3	8.6	6.6
Female	10.6	13.2	6.3	7.3
Age				
Under 15	8.4	11.1	7.4	6.8
15–44	24.2	17.7	14.1	9.1
45–64	26.8	17.6	9.8	11.3
Unknown	44.4	—	33.3	—
Marital status of subscribers and spouses*				
Married subscriber	28.0	16.0	12.6	7.9
Married spouse	17.0	18.4	11.9	14.7
Never married	40.1	18.6	24.6	9.3
Widowed, divorced, separated	25.9	17.9	13.6	7.3
Family income				
Under $10,000	26.6	16.6	16.6	8.9
$10,000–24,999	18.2	16.6	9.6	9.0
$25,000 +	22.8	14.6	15.4	9.4
Unknown	25.0	15.8	12.5	5.3
Occupation of household head				
Faculty	16.8	14.6	13.2	9.5
Other professional staff	20.2	16.8	11.5	9.0
Nonprofessional staff	21.0	16.1	11.2	8.6
Unknown	—	—	—	—
Education of household head				
High school graduate or less	19.7	14.3	9.9	5.0
Some college or college graduate	20.4	15.7	11.0	9.0
Some postgraduate work or graduate degree	20.4	16.0	13.0	9.6
Unknown	50.0	—	25.0	—
Covered by Stanford Major Medical Insurance				
Yes	22.2	15.6	12.4	9.1
No	16.8	18.1	9.7	8.7

TABLE 5. (Continued).

Member Characteristics	Covered Services		Covered Excluding "Free" Services	
	Kaiser	Clinic	Kaiser	Clinic
Covered by outside insurance other than Stanford Major Medical				
Yes	32.5	25.4	15.2	10.1
No	18.8	14.9	11.0	9.0
Satisfaction with the plan				
Completely satisfied	14.4	13.2	7.0	6.9
Generally satisfied	20.5	16.7	11.5	9.9
Not satisfied	31.4	22.0	22.0	14.4
Unknown	61.1	33.3	22.2	33.3
Health status				
Excellent	18.0	14.0	10.8	7.3
Good	22.8	16.2	12.8	10.1
Fair	26.9	33.0	11.5	24.5
Poor	27.3	35.7	9.1	21.4
Unknown	50.0	33.3	—	33.3
Had one or more hospitalizations in 12-month period				
Yes	27.7	40.2	20.0	28.0
No	20.1	14.8	11.2	8.4
Number of in-plan physician and paramedical visits for illness in 12-month period				
None	23.4	13.9	12.7	8.1
1–2	17.1	17.0	8.6	7.8
3–5	19.1	11.2	11.5	7.2
6 or more	22.4	23.2	14.7	15.9
Number of family members covered				
1	34.3	18.1	18.8	8.8
2	26.0	17.4	17.9	11.7
3–4	16.6	17.3	8.8	10.4
5 or more	9.9	9.5	5.2	4.4
Number of years in plan				
1	25.0	20.3	14.3	15.3
2–4	24.4	18.2	15.5	11.9
5–9	15.7	13.6	7.2	6.4
10 or more	—	15.7	—	9.0
Regular source of care				
Specific plan physician	15.3	14.3	9.0	8.0
Plan, no specific physician	19.0	18.0	10.9	11.4
Outside physician	71.2	73.9	57.7	73.9
No regular source of care	55.7	42.5	15.9	10.0

* Dependent children are not included.

TABLE 6. Ordinary Least Squares Estimates of Probability of Using at Least One Out-Of-Plan Covered Physician or Paramedical Service (July 1973 to June 1974)

Independent Variable*	Covered Services		Covered Excluding "Free" Services	
	Kaiser	Clinic	Kaiser	Clinic
Membership status				
Subscriber				
Male	—	—	—	—
Female	.000	−.078	.104	−.017
	(0.0)	(2.55)	(4.60)	(.72)
Spouse				
Male	−.210	−.141	−.047	.000
	(3.64)	(2.19)	(1.01)	(0.0)
Female	−.062	.016	.005	.074
	(2.24)	(.70)	(2.44)	(4.14)
Dependent				
Male	−.130	.027	−.019	.036
	(2.32)	(.48)	(.42)	(.81)
Female	−.136	.030	−.048	.049
	(2.44)	(.54)	(1.07)	(1.11)
Age				
Under 15	−.122	−.073	.021	−.032
	(2.81)	(1.81)	(.60)	(1.00)
15–44	−.024	.000	.038	−.016
	(.94)	(0.0)	(1.88)	(.97)
45+	—	—	—	—
Unknown	.090	−.216	.168	−.125
	(.69)	(1.57)	(1.60)	(1.15)
Marital status†				
Married	—	—	—	—
Never married	.103	.006	.061	−.008
	(2.63)	(.13)	(1.94)	(.22)
Widowed, divorced, separated	.002	.011	−.004	−.018
	(.07)	(.33)	(.16)	(.69)
Income				
Under $10,000	−.037	.023	−.055	−.009
	(.87)	(.60)	(1.63)	(.29)
$10,000–24,999	−.028	.022	−.054	−.009
	(.77)	(1.30)	(1.84)	(.66)
$25,000 +	—	—	—	—
Unknown	.101	−.009	−.012	−.065
	(.98)	(.11)	(.14)	(.99)
Covered by Stanford Major Medical Insurance				
Yes	—	—	—	—
No	.004	.017	−.012	−.017
	(.19)	(.53)	(.79)	(.67)
Covered by outside insurance other than Stanford Major Medical				
Yes	—	—	—	—
No	−.111	−.102	−.031	−.004
	(4.02)	(3.44)	(1.41)	(.17)

TABLE 6. (Continued).

Independent Variable*	Covered Services		Covered Excluding "Free" Services	
	Kaiser	Clinic	Kaiser	Clinic
Satisfaction with the plan				
Completely satisfied	—	—	—	—
Generally satisfied	.057 (2.87)	.036 (2.22)	.035 (2.22)	.031 (2.37)
Not satisfied	.165 (5.36)	.108 (3.01)	.138 (5.57)	.086 (3.04)
Unknown	.312 (3.35)	.102 (.49)	.062 (.83)	.191 (1.17)
Health status				
Excellent or good	—	—	—	—
Fair or poor	.032 (.91)	.144 (4.10)	−.030 (1.04)	.115 (4.17)
Unknown	.306 (1.12)	.146 (.71)	−.123 (.56)	.215 (1.32)
Had one or more in-plan hospitalizations in 12-month period				
Yes	.065 (1.19)	.174 (3.42)	.059 (1.34)	.099 (2.48)
No	—	—	—	—
Number of in-plan physician and paramedical visits for illness in 12-month period	.001 (.52)	.003 (1.29)	.002 (1.24)	.004 (2.73)
Number of family members covered	−.020 (3.11)	−.015 (2.11)	−.016 (3.12)	−.014 (2.42)
Number of years in plan				
Under 5	.047 (2.55)	.047 (2.36)	.048 (3.24)	.062 (3.89)
5+	—	—	—	—
Constant	.373	.241	.118	.774
R^2	.100	.054	.075	.059

* All the independent variables except number of in-plan physician and paramedical visits for illness in 12-month period and number of family members covered are dummy variables equal to 1 if the individual is in the category and equal to 0 otherwise.

† Includes dependents as "never married."

same for widowed, divorced and separated members as for married members.

The relationship between health status as indicated by self-assessment, having had one or more hospitalizations in the 12-month period, and number of in-plan physician and paramedical visits for illness is also not quite the same for the two groups. While the unadjusted figures of Table 5 show an increase in the percentage of out-of-plan users as self-assessed health status worsens for both Kaiser and Clinic plan members, the adjusted figures of Table 6 show that these differences are statistically significant only for Clinic plan members. Clinic plan members who rated their health status as fair or poor were 14.4 per cent more likely to go outside the plan than members who rated their health status as excellent or good. Similarly, Clinic plan members who had had at least one hospitalization in the 12 months of the study were 17.4 per cent more likely to be outside users than members without a hospital episode, while Kaiser plan members showed no significant difference. There appeared to be no relationship, however, for either group, between the number of in-plan physician and paramedical visits for illness and going outside the plan.

It is of interest to note that we found no association between family income and out-of-plan use. The other measures of socioeconomic status, occupation and education, also showed no relationship when we included them in the equations.

Covered Services: Volume of Services

Table 3 shows that the average number of out-of-plan covered services in the 12-month period was 62 per 100 Kaiser and 50 per 100 Clinic plan members. Table 8 presents the average number of out-of-plan covered services per user of such services (first two columns) as well as per 100 plan members (third and fourth columns) by the selected member characteristics used in Table 5. Before examining this table in more detail, however, a look at Table 7 will show at a glance that there were very few heavy out-of-plan users. This table presents the distribution of out-of-plan users of covered services by number of out-of-plan episodes and by number of out-of-plan visits, an episode being defined as one or more physician or paramedical contacts for the same condition or set of conditions. It shows that 76.1 per cent of Kaiser and 77.2 per cent of Clinic plan users had a single episode of out-of-plan care during the 12 months of the study. Similarly, it shows that 48.7 per cent of Kaiser and 50.3 per cent of Clinic plan users had only one out-of-plan physician or paramedical service, and that only 11.5 per cent of Kaiser and 9.8 per cent of Clinic plan outside users had six or more out-of-plan services. Put differently, this means that roughly 12 per cent of Kaiser and 10 per cent of Clinic plan users of out-of-plan covered services had about 50 per cent of all such services. In terms of plan members rather than out-of-plan users, the figures are even more

TABLE 7. Per Cent Distribution of Out-of-Plan Users of Covered Physician and/or Paramedical Services (July 1973 to June 1974)

	Covered Services		Covered Excluding "Free" Services	
	Kaiser	Clinic	Kaiser	Clinic
	100.0% (N = 419)	100.0% (N = 338)	100.0% (N = 237)	100.0% (N = 195)
Number of episodes				
1	76.1	77.2	80.6	82.1
2	17.7	16.6	13.5	14.4
3	3.8	3.8	4.2	2.1
4 or more	2.4	2.4	1.7	1.5
Number of visits				
1	48.7	50.3	51.1	50.8
2	21.2	20.4	19.0	21.0
3	8.4	10.1	8.0	8.7
4	5.5	5.6	5.9	5.1
5	4.8	3.8	5.1	3.1
6 or more	11.5	9.8	11.0	11.3

TABLE 8. Number of Out-of-Plan Covered Physician and/or Paramedical Services (July 1973 to June 1974)

Member Characteristics	Number of Covered Services				Number of Covered Services Excluding "Free" Services			
	Per User		Per 100 Members		Per User		Per 100 Members	
	Kaiser	Clinic	Kaiser	Clinic	Kaiser	Clinic	Kaiser	Clinic
All members	3.05	3.15	61.91	49.74	3.15	3.38	36.20	30.86
Sex								
Male	2.40	2.93	48.82	45.85	2.56	3.51	22.68	26.06
Female	3.67	3.35	74.59	53.40	3.52	3.31	49.28	35.36
Membership status								
Subscriber								
Male	2.24	2.86	61.86	49.92	2.16	3.00	20.51	23.52
Female	4.45	3.44	148.21	50.57	4.14	2.23	98.51	18.49
Spouse								
Male	12.00	22.67	96.00	200.00	20.50	22.67	82.00	200.00
Female	3.11	3.29	56.10	62.47	3.18	3.60	40.70	54.08
Dependent								
Male	2.10	1.90	21.93	25.20	2.13	2.20	18.18	14.59
Female	2.31	3.37	24.52	44.68	1.96	3.43	12.26	24.94
Age								
Under 15	2.22	2.68	18.76	29.64	2.26	3.03	16.70	20.61
15–44	2.96	3.24	71.76	57.37	3.07	3.34	43.24	30.27
45–64	3.74	3.28	100.27	57.92	4.68	3.63	45.89	41.03
Unknown	1.00	—	44.44	—	1.00	—	33.33	—
Marital status of subscribers and spouses*								
Married subscriber	2.89	3.21	81.03	51.48	3.23	3.11	40.75	24.48
Married spouse	3.65	3.95	62.11	72.67	3.93	4.41	46.65	64.80
Never married	3.21	2.56	128.87	47.42	2.63	1.78	64.79	16.49
Widowed, divorced, separated	4.18	2.41	108.16	43.09	4.90	1.67	66.67	12.20
Family income								
Under $10,000	3.48	4.15	92.26	68.79	3.61	1.64	59.96	14.65
$10,000–24,999	2.95	2.73	53.54	45.13	3.09	3.00	29.66	27.01
$25,000 +	1.71	3.62	39.02	53.06	1.53	4.22	23.58	39.50
Unknown	6.00	1.33	150.00	21.05	5.00	2.00	62.50	10.53
Occupation of household head								
Faculty	2.11	3.37	35.33	49.11	1.50	3.83	19.76	36.30
Other professional staff	2.65	2.70	53.61	45.31	2.70	2.90	31.13	25.94
Nonprofessional staff	3.46	3.98	72.67	63.99	3.82	3.62	42.79	31.25
Unknown	—	—	—	—	—	—	—	—
Education of household head								
High school graduate or less	4.40	2.65	86.72	37.86	5.24	2.57	51.61	12.86
Some college or college graduate	2.58	4.53	52.84	71.08	2.30	4.70	25.30	42.16
Some postgraduate work or graduate degree	2.76	2.76	56.32	43.93	2.96	3.02	38.58	28.87
Unknown	2.50	—	125.00	—	4.00	—	100.00	—
Covered by Stanford Major Medical Insurance								
Yes	3.03	3.20	67.23	50.00	3.21	3.44	39.94	31.46
No	3.08	2.56	51.76	46.31	2.99	2.62	29.06	22.82

TABLE 8. (Continued).

Member Characteristics	Number of Covered Services				Number of Covered Services Excluding "Free" Services			
	Per User		Per 100 Members		Per User		Per 100 Members	
	Kaiser	Clinic	Kaiser	Clinic	Kaiser	Clinic	Kaiser	Clinic
Covered by outside insurance other than Stanford Major Medical								
Yes	3.73	3.88	121.10	98.41	4.39	6.11	66.67	61.38
No	2.89	3.03	54.22	45.03	2.93	3.09	32.24	27.90
Satisfaction with the plan								
Completely satisfied	2.61	2.20	37.48	29.09	2.28	2.04	15.94	14.03
Generally satisfied	2.75	3.19	56.29	53.45	2.82	3.48	32.51	34.62
Not satisfied	4.61	6.57	144.62	144.07	4.96	6.88	108.97	99.15
Unknown	3.27	3.00	200.00	100.00	1.50	3.00	33.33	100.00
Health status								
Excellent	2.29	2.96	41.36	41.45	2.28	3.11	24.57	22.58
Good	3.48	3.05	79.41	49.47	3.86	3.48	49.51	35.04
Fair	5.37	4.17	144.62	137.74	5.47	3.69	63.08	90.57
Poor	7.33	5.40	200.00	192.86	16.00	8.00	145.45	171.43
Unknown	3.00	2.00	150.00	66.67	—	2.00	—	66.67
Had one or more hospitalizations in 12-month period								
Yes	3.17	3.12	87.69	125.61	3.00	3.78	60.00	106.10
No	3.04	3.15	61.07	46.72	3.16	3.33	35.42	27.86
Number of in-plan physician and paramedical visits for illness in 12-month period								
None	2.32	2.52	54.23	35.08	2.25	2.69	28.53	21.86
1–2	2.97	3.72	50.69	63.04	3.41	4.22	29.34	33.04
3–5	3.05	3.91	58.23	43.71	3.19	4.03	36.52	28.94
6 or more	4.51	2.45	100.85	56.70	4.23	2.77	62.32	44.13
Number of family members covered								
1	3.29	3.24	112.88	58.77	3.34	2.33	62.45	20.47
2	3.59	3.21	93.18	55.99	3.51	3.76	62.66	44.01
3–4	2.48	3.39	41.12	58.48	2.50	3.89	22.05	40.48
5 or more	2.86	2.00	28.29	19.03	3.41	1.60	17.83	7.08
Number of years in plan								
1	2.91	4.92	72.86	100.00	3.00	1.89	42.86	28.81
2–4	3.00	4.30	73.16	78.38	2.86	4.79	44.32	57.21
5–9	3.15	2.92	49.33	39.63	3.79	3.26	27.29	20.92
10 or more	—	2.58	—	40.42	—	2.81	—	25.21
Regular source of care								
Specific plan physician	3.65	2.41	55.75	34.42	3.63	2.61	32.53	20.97
Plan, no specific physician	2.29	5.18	43.48	93.36	2.43	4.79	26.64	54.50
Outside physician	4.62	10.18	328.85	752.17	4.83	8.27	278.85	647.83
No regular source of care	3.33	3.06	185.23	130.00	2.71	1.25	43.18	12.50

---

* Dependent children are not included.

striking: 2.3 per cent of all Kaiser and 1.5 per cent of all Clinic plan members had roughly 50 per cent of all out-of-plan covered services.

It is therefore not surprising that the data on the number of out-of-plan services *per out-of-plan user* by membership characteristics shown in the first two columns of Table 8 reveal few subgroups who were heavy users of out-of-plan services. The only user groups that stand out as relatively high users are Kaiser plan female subscribers, Kaiser and Clinic plan male spouses, dissatisfied members of both plans, Kaiser and Clinic plan members who rated their health status as poor, Kaiser plan members with six or more in-plan physician or paramedical visits for illness and recent Clinic plan joiners.

The multivariate analysis of the data on out-of-plan users (not reproduced here), where the dependent variable is the number of out-of-plan physician and paramedical services and the independent variables are the member characteristics of out-of-plan users, showed much the same relationships, although not all of them were statistically significant. Dissatisfaction with the plan was found to have a statistically significant relationship with a high use of out-of-plan services. Kaiser plan out-of-plan users who were not satisfied with the plan used 2.02 more outside services than users who were completely satisfied, with a corresponding figure for dissatisfied Clinic plan users of 4.55 more such services. The multivariate analysis also showed that Kaiser and Clinic plan male spouses and Kaiser plan female subscribers who were out-of-plan users used significantly more outside services than male subscribers who used outside services. The figures for spouses are 8.61 more out-of-plan visits for Kaiser and 18.61 more visits for Clinic plan spouses, and 2.19 more visits for Kaiser plan female subscribers.

While the multivariate analysis, like the cross tabulations in Table 8, showed that Kaiser and Clinic plan out-of-plan users who rated their health as fair or poor used more outside services than those who rated it as excellent or good, the differences are not statistically significant. However, the data show that in the case of Kaiser plan out-of-plan users the number of out-of-plan services increased slightly but significantly as the number of in-plan physician and paramedical visits for illness increased; for every additional in-plan visit, the number of out-of-plan services increased by 0.14. Finally, although the multivariate analysis, like the cross tabulations, showed that Clinic plan users who were recent joiners used more outside services than members of longer standing, the difference is not statistically significant.

To summarize, we found few characteristics of out-of-plan users associated with the volume of out-of-plan covered services. Moreover, the data show that a high probability of going outside the plan for covered services does not necessarily go with a high volume of out-of-plan services, and vice versa. For example, while having additional insurance coverage was associated with a higher probability of using some out-of-plan services, it was not associated with a higher volume of outside services. Finally, we found no statistically significant relationship between socioeconomic characteristics—family income, occupation and education—and the volume of out-of-plan services by users, just as we found no relationship between these variables and the probability of out-of-plan use.

We shall not discuss in detail the data on the number of out-of-plan services *per 100 plan members* by member characteristics shown in the third and fourth columns of Table 8. As in the case of the cross tabulations and multivariate analysis of the data on *users* of out-of-plan services, the cross tabulations and multivariate analysis of the data on *plan members* showed much the same relationships between member characteristics and out-of-plan use. Thus, the multivariate analysis (not reproduced here), which duplicates for all members the analysis performed for users of out-of-plan services, showed that for members of both plans, only three characteristics had a statistically significant relationship with the number of out-of-plan services: satisfaction, health status as measured by self-assessment and having additional insurance. Holding other variables constant, dissatisfied Kaiser plan members had 1.09 and dissatisfied Clinic plan members 1.28 more out-of-plan services per year than members who were completely satisfied. Kaiser and Clinic plan members who rated their health status as fair or poor used 0.59 and 0.86 more out-of-plan services, respectively, than members who rated their health as excellent or good. Finally, Kaiser and Clinic members who had no additional insurance used 0.54 and 0.42 fewer out-of-plan services, respectively, than members who had such insurance. In addition, Clinic plan male spouses and Kaiser plan female subscribers were found to be significantly heavier users than male subscribers, with 1.31 and 0.70 more outside services, respectively; and Clinic plan members who had been covered by the plan less than five years also were heavier users than those who had been members longer, with 0.49 more outside services.

Covered Services: Reasons for Out-of-Plan Use

In the telephone interviews all respondents who reported an out-of-plan episode of care for covered services either for themselves or a dependent were asked the reason for having gone to a nonplan provider. The responses are summarized in Table 9, which shows the distribution of outside episodes of care by reason for out-of-plan use. By far the most frequently given reason was preference for a nonplan provider, which was cited as the reason for 50.7 per cent of the Kaiser and 43.5 per cent of the Clinic plan episodes. Needing emergency care was cited as the reason for 13.8 per cent of both the Kaiser and Clinic plan episodes of out-of-plan care. Being too far from plan providers was a reason more often given for Clinic than Kaiser plan outside use, explaining 26.9 per cent and 13.2 per cent of outside episodes, respectively. By contrast, having to wait too long for an

TABLE 9. Per Cent Distribution of Out-of-Plan Episodes of Covered Physician and/or Paramedical Care (July 1973 to June 1974)

Reason for Out-of-Plan Use	Covered Services		Covered Excluding "Free" Services	
	Kaiser	Clinic	Kaiser	Clinic
	100.0% (N = 560)	100.0% (N = 457)	100.0% (N = 303)	100.0% (N = 244)
Needed emergency care	13.8	13.8	10.6	14.3
Too far away	13.2	26.9	16.2	22.5
Too long to get appointment	12.1	6.1	17.2	4.9
Preferred nonplan physician	50.7	43.5	50.2	52.5
Other	10.2	9.6	5.9	5.7

appointment was cited more frequently for Kaiser than Clinic plan episodes, with figures of 12.1 per cent and 6.1 per cent, respectively. Finally, a variety of reasons were given for the remaining 10.2 per cent of Kaiser and 9.6 per cent of Clinic plan episodes of out-of-plan care. Most of the responses in this category indicated confusion on the part of plan members about the coverage status of the service, many of them stating that they believed it was not covered by their plan. We do not know whether plan members who used out-of-plan services because they needed emergency care or were too far away from plan providers were refused reimbursement by the plans (since the plans will pay for such services when they are considered justified) or whether they did not trouble to claim reimbursement.

Out-of-plan users were also asked the condition for which they sought out-of-plan care. About 27 per cent of the Kaiser and 20 per cent of the Clinic plan out-of-plan episodes were for preventive services such as immunizations, check-ups, well-baby care and blood pressure checks. The vast majority of the rest were for minor medical and surgical problems. The most frequently mentioned of these were sore throat, flu, colds, rash, allergy problems, low back pain, and minor injuries and accidents. A very few episodes of out-of-plan care were for consultations about a serious health problem such as cancer.

We also wanted to test the frequently mentioned hypothesis that women members of prepaid plans may have previous ties with a nonplan obstetrician, or mothers with a nonplan pediatrician, whom they continue to consult after joining the prepaid plan. Unfortunately, our data on the fields of specialty of outside providers of covered services are unsatisfactory. Somewhat surprisingly, 60 per cent of Kaiser and 66 per cent of Clinic plan respondents were unable to name the provider's field of specialty. If we assume that they knew if the provider was an obstetrician or pediatrician (and this assumption is probably valid), we can use the data on known specialties for an examination of the validity of the above hypothesis. These data do not support it. Only about one third of outside episodes of covered care of both Kaiser and Clinic plan members involved obstetricians/gynecologists and pediatricians. The rest involved general practitioners, internists, other medical specialists and surgical specialists other than obstetricians/gynecologists.

Noncovered Services

In addition to the use of out-of-plan covered services, we also examined the use of noncovered medical services. Our findings are summarized in Table 10, which shows that 6.7 per cent of Kaiser and 19.7 per cent of Clinic plan members had at least one noncovered service. The relatively high percentage of Clinic plan members who used noncovered services is explained by the fact that routine eye refractions are not covered by the Clinic plan and are therefore counted as noncovered services. When Clinic plan members whose only noncovered service was such an eye examination are excluded, the proportion who used noncovered services becomes much the same as that of Kaiser plan members, 6.0 per cent. The average number of out-of-plan services per user was 15.15 for Kaiser and 5.85 for Clinic plan members; the latter figure becomes 16.22 when eye refractions are excluded.§§ In terms of noncovered services per member, the figures are 1.02 for Kaiser and 1.15 for Clinic plan members (or 0.98 for the latter when eye examinations are excluded).

---

§§ These figures reflect the fact that relatively many Clinic plan members had eye examinations, but few if any had more than one such service during the 12-month period; by contrast, relatively few members of either plan had other noncovered services, but those who did had relatively large numbers of such services.

TABLE 10. Out-of-Plan Use of Noncovered Physician and/or Paramedical Services (July 1973 to June 1974)

	Kaiser	Clinic	
% with at least one noncovered service	6.7	19.7	(6.0)*
% with at least one noncovered psychiatric service	2.8	4.5	
% with at least one noncovered eye examination	—	14.8	
Mean no. of noncovered services per user	15.15	5.85	(16.22)†
Mean no. of noncovered services per member	1.02	1.15	(.98)†
Mean no. of noncovered psychiatric services per member	.64	.78	
Mean no. of noncovered services rendered by paramedical personnel per member	.48	.32	(.30)†

* Excluding those with routine eye examinations only.

† Excluding routine eye examinations.

The principal noncovered service reported by members of both plans was psychiatric care. Almost 3 per cent of Kaiser and 4.5 per cent of Clinic plan members had at least one psychiatric service during the 1-year period of the study. The average number of psychiatric services per plan member was 0.64 for Kaiser and 0.78 for Clinic plan members, thus representing about two thirds of all noncovered services of members of both plans. For Clinic plan members, routine eye refractions were another fairly frequently used service, accounting for about 15 per cent of noncovered services. It is of interest to note that Kaiser plan members relied for almost half of their noncovered services on nonphysician personnel, while Clinic plan members tended to consult physicians and obtained only just over one quarter of their noncovered services from such personnel. The most frequently consulted nonphysicians were chiropractors, psychologists and—in the case of Clinic plan members—optometrists.

Out-of-Pocket Expenses for Covered and Noncovered Services

Out-of-pocket expenditures for medical services obtained outside the plan are shown in Table 11.[|||] Both the distributions of such expenditures and average expenditures per user of out-of-plan services and per plan member are given separately for covered services, noncovered services and both types combined. Total out-of-pocket expenses for out-of-plan covered and/or noncovered services averaged $89.83 for Kaiser and $114.64 for Clinic plan users of such services, or $22.66 per Kaiser and $36.39 per Clinic plan member. As the various figures show, members of both plans spent substantially less for covered than for noncovered services, expenditures for covered services representing 28 per cent of the total out-of-pocket expenditures of Kaiser plan members and 25 per cent of those of Clinic plan members.

[|||] The figures in Table 11 understate out-of-pocket expenses to an unknown degree because they contain a small number of members of both plans who said they had had some out-of-pocket expenses for covered and/or noncovered services but did not remember the amount. If we assume that such persons had the same average out-of-pocket expenses as users of out-of-plan services who reported having had some expenses and remembered how much these had been, we get the following estimates of out-of-pocket expenses per plan member:

	Covered services	*Noncovered services*	*All out-of-plan services*
Kaiser	$6.75	$16.84	$23.59
Clinic	9.59	29.51	39.10

These estimates may actually be on the high side since they are based on expenses by out-of-plan users with expenses, some of whom had high expenses; however, it is reasonable to assume that those who reported to have made a payment but did not remember the amount very probably had only minor expenses.

TABLE 11. Distribution of Kaiser and Clinic Plan Out-of Plan Users by Amount of Out-of-Pocket Expenses (July 1973 to June 1974)

Range of Out-of-Pocket Expenses	Covered Services		Noncovered Services		All Out-of-Plan Services	
	Kaiser	Clinic	Kaiser	Clinic	Kaiser	Clinic
	100.0% (N = 419)	100.0% (N = 338)	100.0% (N = 138)	100.0% (N = 422)	100.0% (N = 520)	100.0% (N = 679)
$0	47.5	45.0	5.8	9.0	37.3	22.7
$1–49	35.8	38.8	35.5	65.8	35.0	53.3
$50–99	11.0	5.3	6.5	3.8	9.6	5.4
$100–499	4.9	8.9	39.8	14.5	14.2	13.0
$500 +	.7	2.1	12.3	6.9	3.8	5.6
Mean expense						
per user	$31.11	$57.69	$244.05	$138.25	$89.83	$114.64
per member	6.32	9.12	16.34	27.27	22.66	36.39

The distributions of out-of-plan users by amount of out-of-pocket expenses show that a relatively small percentage of all users had high expenses. Only 3.8 per cent of Kaiser and 5.6 per cent of Clinic plan users of out-of-plan covered and noncovered services had out-of-pocket expenses of $500 or more, and only 18.0 per cent of Kaiser and 18.6 per cent of Clinic plan users had expenses of $100 or more. Put differently, this means that 18.0 per cent of Kaiser plan out-of-plan users, or 4.6 per cent of all Kaiser plan members, were responsible for 82.8 per cent of total out-of-pocket expenditures for covered and noncovered services. Similarly, 18.6 per cent of Clinic plan out-of-plan users, or 5.9 per cent of all Clinic plan members, accounted for 84.8 per cent of total out-of-pocket expenses.

The distributions of out-of-plan users of covered services by amount of out-of-pocket expenditures show the same picture. Only 5.7 per cent of Kaiser and 11.0 per cent of Clinic plan users of such services spent $100 or more for them, and only 16.7 per cent and 16.3 per cent, respectively, spent $50 or more. In terms of the percentage of total expenditures for out-of-plan covered services, this means that 3.4 per cent of Kaiser plan members accounted for just over 70 per cent of out-of-pocket expenses, and 2.6 per cent of Clinic plan members for a little over 80 per cent. The distributions of out-of-plan users of noncovered services are somewhat less skewed. Nevertheless, again a very small percentage of all plan members accounted for a high percentage of total out-of-pocket expenses for these services, 3.5 per cent of Kaiser plan members accounting for 94 per cent, and 4.2 per cent of Clinic plan members for 86 per cent of such expenses. Like the data on the number of out-of-plan noncovered services by type of service, the expenditure data show that psychiatric care was the principal service obtained, accounting for 72 per cent of Kaiser and 76 per cent of Clinic plan members' expenditures for noncovered services.

We have made an estimate of out-of-pocket expenses for out-of-plan services as a percentage of the total costs of the two plans, i.e., expenditures for premiums (both the share paid by plan members and that contributed by Stanford University), plus the Kaiser copayment of $1 for office visits and the Clinic 25 per cent coinsurance payment, plus out-of-pocket expenses for out-of-plan care.[##] In the case of Kaiser plan members, expenditures for all out-of-plan services represented 12.3 per cent of total costs and expenditures for covered services alone 3.4 per cent. The corresponding figures for Clinic plan members are 13.3 per cent and 3.3 per cent of total costs, respectively. Clearly, expenditures for services that members of both plans could have obtained from plan providers are a minor fraction of total costs.

Comparison With Other Studies

Extent of Out-of-Plan Use of Covered Services

In view of the lack of comparability of data on out-of-plan use discussed earlier, it is difficult to assess how the extent of out-of-plan use of services, especially of covered services, that we found in our study compares with that reported for members of other prepaid plans. The only study with data comparable to ours, in the sense that a distinction is made between covered and noncovered services and attention is focused on covered services, is that by Pope et al.[18] for the Kaiser-Oregon plan in 1971. They found that between 6 per cent and 7 per cent of a sample of 2,500 adults who had been members of the Kaiser plan for at least two years had used nonplan physicians for covered services; the lower figure refers to members who had to pay for such services, the higher includes members who also received "free" services. In addition, the authors found that the mean number of conditions for which outside services were used was only slightly above one, and that most episodes of out-of-plan care involved only one visit. Even if we make the rather liberal assumption, based on this information, that out-of-plan users averaged 1.5 outside covered physician services per year, this translates into only 10–11 such services per 100 members per year, very much less than the volume reported by the Kaiser and Clinic plan members in our study.

Data from three other studies conducted in the late 1960s and early 1970s, although not strictly comparable to our data because they refer to *all* out-of-plan physician services, covered and noncovered, also suggest that out-of-plan use of covered services of our two groups is relatively high. Richardson et al.,[19] in their 1971–1974 study of a near-poverty population enrolled in a large prepaid group practice plan in Seattle (Group Health Cooperative of Puget Sound), found that 16–19 per cent of enrollees had at least one visit to an outside physician and that 11 per cent of all their physician visits, in-plan and out-of-plan, were to an outside source. Hetherington et al.,[16] reporting on out-of-plan use by members of one large and one small prepaid group practice plan in the late 1960s, found that 10.5 per cent of all physician visits by members of the large group and 6.9 per cent of those by members of the small group were to out-of-plan physicians. In a study of three HMOs in Rochester, New York, Ashcraft et al.[5] report that out-of-plan use of ambulatory services was minimal, with only 3 per cent of all ambulatory services provided outside the plans. A fourth study, of out-of-plan use by members of Kaiser-Northern California in 1971,[22] reported that 35 per cent of the membership had used non-Kaiser doctors at some time in the preceding two years, which makes the data difficult to compare with ours. In addition, 22 per cent of these out-of-plan users were later reimbursed by Kaiser be-

---

[##] The figures below show mean costs of the two plans per plan member by type of cost, and mean out-of-pocket expenditures (i.e., costs excluding the share of the premium paid by Stanford University):

	Kaiser	*Clinic*
Mean costs per member	*$184.29*	*$273.53*
Premium	158.61	206.16
Copayment	3.01	30.98
Out-of-plan expenses	22.66	36.39
Mean out-of-pocket expenses per member	$ 95.65	*$191.44*

cause their out-of-plan services were emergency visits while out of the Kaiser service area.

Three studies dealing with Medicaid[12,15] and Medicare[13] enrollees in prepaid practice plans and carried out in the early 1970s found out-of-plan use of all types ranging from practically zero[12] to almost two thirds of all physician services obtained by the enrollees.[15] Because of the special characteristics of these groups, however, these data are not comparable to ours.

It is of interest to note that the few early studies of the utilization experience of prepaid plans conducted in the 1950s show relatively high rates of out-of-plan use. A study of Kaiser-Northern California members,[7] referring to all out-of-plan use in 1958, found that one third of the families in the sample reported one or more visits to physicians outside the plan. It was also found that plan members received 72 per cent of their office and home visits from plan providers, 12 per cent from outside physicians at their own expense and 15 per cent from outside providers at no charge. Another study, reporting on the experience of a St. Louis prepaid group practice plan in the early 1950s and also referring to all out-of-plan services,[21] found that almost half the families in the sample had used some outside physician services during the year of the study and that plan members had received 23 per cent of their total professional services from out-of-plan providers. A partial explanation of the high percentage of out-of-plan users is the fact that the figure refers to families, not individuals, which means that a multimember family is counted as an out-of-plan user even if only one member used out-of-plan services. An explanation of the high proportion of services obtained from out-of-plan providers (making a reasonable allowance for noncovered services) may be that the plans were fairly new so that many plan members still had ties to a previous family physician. Another possible reason is that the plans were not only new but a novel form of health care delivery that took some getting used to. There is also some evidence that a large share of the out-of-plan physician visits were emergency or home visits when plan members found it more convenient to consult a neighborhood physician rather than go to the plan physician or facility.

In summary, there is some evidence that Kaiser and Clinic plan members may have been somewhat heavier users of out-of-plan covered services than members of other prepaid plans in the 1970s, though the evidence is not conclusive because most other studies deal with all out-of-plan services rather than only covered services. One possible explanation may be the very high availability of other sources of care in the Palo Alto–Stanford area and the relative medical sophistication of our groups that this wealth in medical resources is likely to entail. It is also possible, however, that other studies have underestimated out-of-plan use. The Kaiser-Oregon study and, so far as we can judge, most other studies as well, used a 1-year recall period for information on out-of-plan use, while we contacted plan members every three months. In addition, our initial household interview included questions on other insurance coverage such as student health policies and eligibility for Veterans Administration benefits and CHAMPUS, and our interviewers probed for use of such services when contacting plan members who had reported such coverage. Despite these frequent and probing contacts, we found that a few enrollees who had used out-of-plan services could not remember what they had paid for them. It is quite likely, therefore, that a 1-year recall period is too long for remembering such a relatively rare and usually inexpensive event as use of an out-of-plan service and thus leads to an underestimate of out-of-plan use.

Characteristics of Out-of-Plan Users of Covered Services

Most of the studies of out-of-plan use make at least some attempt to examine member characteristics associated with going outside the plan, usually in terms of the percentage of plan members with some out-of-plan services, but occasionally in terms of the percentage of medical services obtained outside the plan. Some of their findings agree with ours, others disagree. Not surprisingly, dissatisfaction with the plan is generally found to be an important factor. The studies by Pope et al.,[18] Freidson,[11] Fuller et al.,[12] Anderson and Eggers,[4] and Franklin and McLemore[10] all found that dissatisfied plan members were more likely to use out-of-plan services and/or to obtain a substantial percentage of their care outside the plan than satisfied members. The study by Pope et al. of Kaiser-Oregon plan members provides the most careful analysis of specific reasons for dissatisfaction. The most important of these are complaints about the quality of care, difficulty of access to physicians and impersonality of care.

The evidence on the relationship between health status and out-of-plan use is mixed. Pope et al.[18] found no relationship in their study of Kaiser-Oregon plan members. By contrast, Smillie[22] reported that Kaiser-Northern California plan members who were heavy users of in-plan services (presumably because they were ill, though he does not say so) were the ones who were the most likely to use outside services. Similarly, Anderson and Eggers[4] found that students who experienced disability were more likely to see physicians outside their prepaid plan than healthier enrollees. The study by Greenfield et al.[13] of Medicare members enrolled in HIP found an association between frequency of out-of-plan use and severity of illness as measured by frequency of hospitalization, average amount of reimbursed charges for hospitalization, average number of days in hospital for persons with at least one hospital episode, frequency of surgical hospitalization and mortality. The values for all of these measures were found to be greater for out-of-plan users than for those who used only HIP services. On balance, despite the negative findings of the Kaiser-Oregon study, it can probably be concluded from the other evidence that plan members in poor health are more likely to go outside the plan than members in good health, but that the evidence that they will also use more outside services is slim.

Some of the studies examine the role of length of plan membership, but the results to date are unsatisfactory. Freidson,[11] Anderson and Eggers,[4] and Franklin and McLemore[10] all found that the probability of out-of-plan use increases with length of time in the plan. However, what they measure is out-of-plan use *at any time* during the enrollee's plan membership. This makes their findings merely what common sense suggests, i.e., that the longer a person is enrolled in a plan, the more likely it becomes that he/she will go outside it *at some time* over the course of the years. Simon and Rubushka,[21] in their 1956 study, found some indirect evidence that the probability of a plan member being an out-of-plan user in a given period of time declines with length of plan membership because "As the members begin to use the plan they find less need to seek services elsewhere." Nevertheless, our finding of a small but significant inverse relationship between out-of-plan use and length of plan membership must be regarded as tentative until confirmed by other studies, especially since our evidence comes primarily from just one of the plans, the Clinic plan.

Several studies look at the roles of age, sex and marital status. Except for the finding that children are the least likely to go outside the plan, on which the two studies that examine this factor[6,18] agree with each other and with our findings, there is no general agreement about the relationship between these characteristics and out-of-plan use. The groups for which one study or another found a high probability of outside use include women compared with men,[18] married women compared with single men and women and married men,[6] and married students—male and female—compared with single students.[4] Since our study also found no consistent relationship between these variables and out-of-plan use for members of the two plans, it is possible that there is no generalizable relationship and that factors other than age, sex and marital status are more important. At least one study, however, confirms our finding that family size is inversely related to the likelihood of out-of-plan use. Pope et al.[18] found that Kaiser-Oregon plan members who were outside users ". . . are more likely than others to be in one or two person families (59% compared to 50%). Or, to put it another way, outside users are more likely (57%) than others (49%) to be in Health Plan units with no children." Despite this confirmation of our findings, further studies are needed before definite conclusions about the relationship between family size and outside use can be made.

While we found no association between outside use and socioeconomic status as measured by family income, occupation and education of the head of the household, two studies[11,18] found that the higher socioeconomic groups were more likely than others to go outside the plan. For example, the Kaiser-Oregon study[18] shows that "About 37% of the outside users compared to 30% of all others were in the highest educational category; 28% vs. 22% had an income of $15,000 or more; 43% vs. 35% were in white-collar occupations." Thus, this is another area where further investigation is needed before definite conclusions can be drawn.

None of the studies we examined or know of mentions the availability of additional insurance coverage and its possible association with out-of-plan use. This is rather surprising, at least in the case of the more recent studies carried out at a time when duplicate coverage was no longer a rare exception. Since we found that such additional insurance coverage was associated with a significantly higher probability of out-of-plan use, and since the increase in the number of married women working outside the home is very probably leading to an increase in the percentage of the population with duplicate insurance coverage, this is a factor that needs careful scrutiny.

Conclusion

In conclusion, a few observations on our findings deserve special emphasis. First, our study found that while a fairly high percentage of members of both plans used some out-of-plan covered services, the vast majority of these out-of-plan users were occasional users, and only a very small number were really heavy users, accounting for most of the out-of-plan services of the two groups. Since at least two other studies[11,18] cite evidence that confirms this finding, it may well be true of prepaid plans in general, although future studies should pay more attention than have most studies in the past to the distribution of out-of-plan users by amount of out-of-plan services or expenditures. If future studies also confirm our finding, it becomes rather misleading to measure out-of-plan use solely in terms of the percentage of the membership with some out-of-plan use or to express out-of-plan use in terms of the average number of out-of-plan services or the average amount of expenditures for out-of-plan services per plan member. Occasional out-of-plan use is probably a fact of life that will occur in all prepaid plans and is of relatively minor importance. It is the high out-of-plan users who deserve special attention.

Second, while our study was able to determine some of the member characteristics associated with high out-of-plan use of covered services, it could not answer the important question of whether the high out-of-plan users are a hard core of members who year after year obtain a substantial portion of their medical care from nonplan providers, or whether high users in one year drop out of the plan after a while and are replaced by other high out-of-plan users. This piece of information is essential for a better understanding of out-of-plan use and should be explored in future studies.

Third, our study and several other studies of the subject showed that dissatisfaction is a principal (if not *the* principal) characteristic associated with high out-of-plan use. Few studies, including ours, have examined what exactly is at the root of such dissatisfaction. Thus, this is another question that deserves more attention in future studies. Furthermore, if it should be found that dissatisfied members who use nonplan providers for a large part of their medical needs remain members of the plans year after year, the question arises as to why they stay with the plan.

Finally, and perhaps most important, there is the question of duplicate insurance coverage. As pointed out above, such duplicate coverage is becoming increasingly common and is likely to continue to grow in the future. This trend needs careful study, not only because it is likely to affect out-of-plan use under prepaid and other health insurance plans, but—and this is more important—because it complicates the evaluation of medical and hospital utilization rates under such plans. A person covered by two plans may, for example, obtain half of his/her medical care from one plan and the other half from the other. This lowers utilization rates under both plans, which in turn may lead to undeserved criticism of both plans for not providing sufficient ambulatory care and/or to undeserved praise for reducing the use of hospital care. Luft,[23] for example, has shown that about half of the decline in the Blue Cross hospital admission rate nationally during the period from 1968 to 1978 is attributable to increasing duplicate coverage. Clearly, this is an area in need of further study.

References

1. Scitovsky AA, McCall N, Benham L. Factors affecting the choice between two prepaid plans. Med Care 1978;16:660.

2. Scitovsky AA, Benham L, McCall N. Use of physician services under two prepaid plans. Med Care 1979;17:441.

3. Scitovsky AA, McCall N. Use of hospital services under two prepaid plans. Med Care 1980;18:30.

4. Anderson JG, Eggers PW. Patterns of utilization among members of a prepaid group practice. Inquiry 1976;XIII:382.

5. Ashcraft M, Penchansky R, Berki SE, Fortus RS, Gray J. Expectations and experience of HMO enrollees after one year: an analysis of satisfaction, utilization, and costs. Med Care 1978;16:14.

6. Bashshur RL, Metzner CA, Worden C. Consumer satisfaction with group practice, the CHA case. Am J Pub Health 1967;57:1991.

7. Columbia University School of Public Health and Administrative Medicine. Family medical care under three types of health insurance. New York: Foundation of Employee Health, Medical Care and Welfare, Inc., 1962.

8. Corbin M, Krute A. Some aspects of Medicare experience with group-practice prepayment plans. Soc Secur Bull 1975;38:3.

9. Donabedian A. An evaluation of prepaid group practice. Inquiry 1969;VI(3):3.

10. Franklin BJ, McLemore SD. Factors affecting the choice of medical care among university students. J Health Soc Behav 1970;11:311.

11. Freidson E. The organization of medical practice and patient behavior. Am J Public Health 1961;51:43.

12. Fuller NA, Patera M. Report of a study of Medicaid utilization of services in a prepaid group practice health plan. DHEW, Public Health Service, Bureau of Medical Services, January 1976.

13. Greenfield CA, Densen PM, Jones EW, Kleinman JC, Miller JA. Use of out-of-plan services by Medicare members of HIP. Health Serv Res 1978; 13:243.

14. Harris L, et al. American attitudes toward health maintenance organizations. New York: July

1980. (Prepared for The Henry J. Kaiser Family Foundation.)

15. Hester J, Sussman E. Medicaid prepayment: concept and implementation. Milbank Mem Fund Q 1974;52:415.

16. Hetherington RW, Hopkins DE, Roemer MI. Health insurance plans: promise and performance. New York: Wiley-Interscience, 1975.

17. Luft HS. Health maintenance organizations: dimensions of performance. New York: Wiley-Interscience, 1981.

18. Pope CR, Freeborn DK, Greenlick MR. Use of outside physicians by members of a group practice prepayment plan. Presented at A.P.H.A. Annual Meeting, November 1972.

19. Richardson WC, Boscha MV, Weaver BL, Drucker WL, Diehr PK. The Seattle prepaid health care project: comparison of health services delivery. University of Washington: Dept. of Health Services, School of Public Health and Community Medicine, November 1976.

20. Roemer MI, Hetherington RW, Hopkins CE, Gerst AE, Parsons E, Long DM. Health insurance effects: services, expenditures, and attitudes under three types of plans. Bureau of Public Health Economics, Research Series No. 16, School of Public Health, University of Michigan, 1972.

21. Simon N, Rubushka S. Membership attitudes in the Labor Health Institute of St. Louis. Am J Public Health 1956;46:716.

22. Smillie JG. Testimony before the United States House of Representatives, Subcommittee on Public Health and Environment of the Committee on Interstate and Foreign Commerce, Health Maintenance Organizations. Hearings, April 11-13, 1972. Serial 92-88.

23. Luft HS. Diverging trends in hospital utilization: fact or artifact? Forthcoming in Med Care.

MEDICAL CARE
December 1981, Vol. XIX, No. 12

Membership Duration and Utilization Rates in a Prepaid Group Practice

MARY JANE GRIFFITH, PH.D.,* AND NICHOLAS BALOFF, PH.D.†

The relationship between utilization rates and duration of membership in prepaid group practices can have important implications for national policy and HMO administration. The results of this study demonstrate a substantial reduction in utilization rates with increasing duration of membership over a 5-year period in a developing prepaid group practice (PGP) located in urban St. Louis. This "start-up effect" is strongly evident across different socioeconomic subgroupings of the study population (sex, race, income, education and age) and for different employers and cohorts. The implications of widespread start-ups for health care financing policy and for important PGP management decisions are discussed, and suggestions are made for future study of the start-up phenomenon.

AN IMPORTANT AND TIMELY TOPIC is that of the effect of duration of enrollment on rates of utilization of health care services in prepaid group practices. Results of investigations to date are mixed. Some authors have reported the existence of a definite "start-up effect" that is marked by a substantial reduction in the utilization rates of new enrollees as their length of membership in PGPs increases; this phenomenon appears to be the result of initially high rates of utilization by new members as they explore the plan, test the benefits offered, and are screened and treated for preexisting illnesses, followed by lower utilization rates as members become acquainted with the plan and their illnesses are treated.[1,2] On the other hand, a more recent study concluded that no such start-up effect exists.[3]

It is important to resolve this issue because the existence, generality and magnitude of such start-ups can have important implications for national policy and for the effective administration of PGPs. It could be proposed, for example, that the federal government encourage HMO development generally by providing extra reimbursement for the start-up costs associated with enrolling new members.

The government might also provide special incentives for enrolling groups of poor or medically disadvantaged individuals in PGPs. Administrators could fear that such individuals would have a high degree of unmet health needs that would result in high initial rates of utilization and a drama-

---

* Assistant Professor, Health Administration and Planning Program, Washington University School of Medicine, St. Louis, Missouri.

† Professor, Graduate School of Business Administration, Washington University, St. Louis, Missouri.

Support for this research was provided by the W. K. Kellogg Foundation.

Address for reprints: Mary Jane Griffith, Health Administration and Planning Program, Washington University School of Medicine, 724 South Euclid Avenue, St. Louis, MO 63110.

0025-7079/81/1200/1194/$01.35 © J. B. Lippincott Co.

tic start-up effect. The cost of these start-ups could therefore be an appreciable deterrent to the enrollment of such groups unless special financing mechanisms were developed.

Substantial start-ups also have many important consequences for PGP management because large changes in utilization affect the "service capacity" of a program, which, in turn, alters the essential bases for most planning and control functions within the organization. For example, a plan's marketing strategy should be determined in relation to the start-up effect; and administrators should also reflect this effect in their staffing and equipment acquisition decisions, premium calculations and facility expansion plans.

Past Studies

Past studies provide only equivocal support for the existence of a general start-up phenomenon. Among older studies, four showed some decline in the rate of utilization with increases in membership duration.[4-7] The strength and implications of these results are diminished somewhat, however; since these studies typically relied on cross-sectional—rather than longitudinal—analyses, intervening variables were often not controlled, and other methodological difficulties were encountered.

Two articles recently published in this journal speak most directly to the existence of the start-up phenomenon. In the first of these, Yesalis and Bonnet reported the existence of a substantial start-up effect based upon their longitudinal analysis of new enrollees in a PGP located in Columbia, Maryland.[2] They found that three indices of utilization—the total number of visits to providers, the proportion of well-care visits and the rate of new prescriptions filled—decreased by as much as 50–60 per cent over a time period of roughly two years. The authors described their study population as being homogeneously young, educated and upper-middle class. They did not, therefore, examine the impact of age, sex or socioeconomic variations on the start-up effect.

These findings are questioned in the most recent longitudinal study of the start-up effect. Based upon their analyses of utilization data drawn from the Kaiser-Permanente Health Care Plan in Portland, Mullooly and Freeborn concluded that "in general, . . . the length of time on the plan does not appear to be a major factor in accounting for changes in ambulatory care utilization or in the amount of services used."[3]‡ Furthermore, they critically reviewed Yesalis and Bonnet's findings and suggested that the start-up effect in this study was a statistical artifact of the methodology employed: "the start-up utilization effects reported by Yesalis and Bonnet became apparent only after statistical techniques adjusting for seasonality, changes in the ratio of provider manpower, health plan hours and benefit packages were made."[3]

The results of Mullooly and Freeborn's study bear closer examination. They relate to two different groups drawn from the Kaiser-Permanente Plan for the period of 1968 through 1972. The first study group included a sample of persons normally enrolled in the Kaiser Health Plan, but the second study group was an experimental population drawn from participants in an OEO Neighborhood Health Center Project. Utilization of ambulatory care services by these two study populations was analyzed for a period of approximately 5 years (21 quarters). The two study populations were divided into subgroups by age and sex categories, but no specific analysis by income, education level or race was performed.

---

‡ Another paper that used essentially the same data base and methodology as Mullooly and Freeborn also reported finding no evidence to support the start-up phenomenon.[8]

Mullooly and Freeborn's general conclusion about the absence of a start-up effect was based importantly upon their graphic analyses of utilization rates over time for the two different study populations taken in the aggregate and the separate graphic plots for several different sex–age subgroups of the two populations. Our interpretation of these graphic results is somewhat different from theirs. Although the two aggregate plots show no start-up and are therefore consistent with their conclusion, the plots for several sex–age subgroups seem to show definite and pronounced start-ups. Specifically, the plots for males and females aged 0 through 14 years in both study populations show a rather clear *downward trend* in utilization; in all four of these cases, there were sizable reductions in utilization rates for approximately three years.[§] Older members of the two study populations appear to have had quite different utilization histories. The utilization plots for both males and females aged 45 through 59 years in both populations show sizable *upward* trends in utilization occurring during the 21 quarters of the study.[‖]

These detailed results for the subgroups raise questions about the meaning of the aggregate plots. Because of the opposite and potentially canceling utilization trends of the subgroups shown in the study, the aggregate utilization histories and the conclusions drawn from them may have limited generality and application.[#] For example, generalizations or extensions of the conclusions to other plans may be restricted by the specific sizes and mix of the sex–age subgroups of these plans. The applicability of these results may also be limited for both managerial and policy purposes. Managers need to know the utilization behavior of specific socioeconomic subgroups in order to plan and operate effectively; average results for an entire population, for example, may not indicate the different and nonsubstitutable services that will be demanded by different subgroups (such as pediatrics versus geriatrics). In policy considerations as well, attention is often focused on specific socioeconomic groups, rather than on the total population.

In summary, the past literature on the subject of the start-up effect seems to be incomplete in several ways: It provides little consistent evidence for the existence of a start-up effect; the effect of different socioeconomic status (SES) classifications upon utilization behavior has not been studied sufficiently; too little attention has been paid to important policy and management implications of start-ups.

These three issues are examined below: we present evidence of a strong start-up effect in a plan that resembles many young PGPs that have developed in the past decade. Our analyses explicitly recognize variations in the characteristics of the start-up effect for different SES groups based upon sex, race, income, education and age—and for different employers and cohorts. We conclude by discussing important policy and managerial implications of the results.

[§] Starting in the second quarter of membership and continuing through the 12th quarter, the number of provider visits per quarter were decreased by approximately 30–40 per cent in three of the four plots, and by approximately 25 per cent in the fourth.

[‖] The two female subgroups show an upward trend throughout the 5-year history, yielding overall increases in utilization of approximately 20 per cent and 30 per cent in the two cases. For males aged 45 through 59 years, this upward trend develops later, but the increases in utilization from the 12th to the 21st quarter are approximately 20 per cent and 40 per cent.

[#] The utilization plots for the other sex–age groups are not given, so we cannot evaluate their effect on the aggregate. Based upon the evidence shown, it seems quite possible that the aggregate graphs are artifacts of cumulating different utilization trends for the subgroups, resulting in an aggregate curve that does not describe its components well.[9]

Study Setting

The setting of this study is the Medical Care Group (MCG), a closed-panel prepaid group practice located in St. Louis, Missouri. The plan began in 1969 and has grown to a current enrollment of 24,000. Plan members normally enroll through their employers, who contract with MCG for services. The large number of employers (160), of different size and nature (business, government, education, etc.), has resulted in a diverse membership.

Primary care services are delivered at MCG's ambulatory care facility, which is located in an urban area next to a large number of acute care facilities and major teaching centers.[10] All patients are assigned a primary care physician with whom they normally interact. Fourteen physicians, three physician's assistants, 16 nurses and four nurse practitioners are currently employed. MCG contracts with the Washington University School of Medicine for consultations with specialists, surgical procedures, laboratory tests, X-rays and other specialized procedures and services.

MCG has always offered a comprehensive package of ambulatory, hospital and home services to its enrollees without deductibles.

Data and Analysis

The utilization data upon which this study is based were drawn from the outpatient encounter file at MCG for the 5-year period from 1973 to 1977. This file includes all outpatient encounters and provides information relating to the purpose of the encounter, diagnosis, procedures performed and attending provider.

The socioeconomic data used in the study were collected through a mail survey of enrollees that was conducted by MCG in 1978. Only respondents to this survey (63 per cent of the total MCG population) who enrolled after January 1, 1973 and continued membership through 1977 are included in the study population.[¶]

The utilization measure used for analysis is "provider visits," which includes all encounters with either a physician or a physician's assistant that took place at the ambulatory care facility. Encounters with other personnel (nurse, social worker, etc.) and in other settings (emergency room, patient's home, etc.) are excluded. The number of provider visits made in each quarter since a member's enrollment are averaged over all enrollees for each quarter and converted to an annual rate. The "provider visit" measure was defined in this way to make it comparable to those used in the studies by Yesalis and Bonnet and Mullooly and Freeborn.[**]

Duration of membership was measured in relation to "calendar quarters" of enrollment in the plan. That is, the aggregated number of provider visits was compared with the number of calendar quarters since a person's enrollment.[††]

[¶] The sex, family size and employment characteristics of the nonrespondents are similar to those of the study population.[11] Analysis without socioeconomic subcategorization of these nonrespondents showed utilization patterns similar to those of the study population on all important characteristics, but the overall level of utilization was slightly lower. Persons who disenrolled from MCG (two to five per cent per year) are also excluded from the analyses.

[**] It is not identical, however, because of differences between the plans in the way providers are used. For example, we have excluded nurses—whom Mullooly and Freeborn included—because they have a more managerial or mediating role at MCG.

[††] The calendar quarter definition of duration is consistent with the precedent set by Yesalis and Bonnet. It differs from that used by Mullooly and Freeborn, who defined the first quarter of membership to be the first 90 days of membership, regardless of the calendar date of enrollment. The calendar quarter approach is more useful for two reasons: 1) it relates to the normal data aggregation base of PGPs and is therefore more likely to be chosen in future research, while the other definition would require costly reorganization of data files; and 2) calendar quarters are also the normal basis for most of the administrative functions affected by start-ups. The calendar quarter definition has the drawback, however, of including in the same quarter individuals of differing exact durations of membership. The implica-

TABLE 1. Annualized Number of Provider Visits by Quarter of Enrollment

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Population(1)	3.48 (10730)	5.37 (9851)	4.88 (8472)	4.36 (7555)	4.55 (5939)	4.55 (5290)	4.47 (4524)	3.89 (4361)	3.85 (2274)	4.13 (1847)	4.01 (1632)	3.91 (1548)	3.61 (1144)	3.82 (806)	3.71 (606)	3.54 (570)	3.65 (301)	4.10 (278)	3.23 (259)	3.14 (98)
Sex: Female(2)	3.76 (5613)	5.86 (5154)	5.52 (4456)	4.94 (3973)	5.13 (3116)	5.27 (2761)	5.11 (2341)	4.47 (2244)	4.32 (1174)	4.80 (953)	4.49 (845)	4.68 (797)	4.25 (585)	4.54 (430)	4.84 (319)	4.65 (301)	4.52 (153)	4.61 (143)	3.55 (134)	4.08 (50)
Male(3)	3.17 (5117)	4.82 (4697)	4.18 (4016)	3.72 (3582)	3.91 (2823)	3.77 (2529)	3.78 (2183)	3.27 (2117)	3.35 (1100)	3.41 (894)	3.51 (787)	3.09 (751)	2.94 (559)	3.00 (376)	2.98 (287)	2.35 (269)	2.76 (148)	3.55 (135)	2.88 (125)	2.17 (48)
Race: White(4)	3.59 (8070)	5.52 (7430)	4.93 (6545)	4.45 (5782)	4.63 (4678)	4.57 (4126)	4.51 (3483)	3.85 (3359)	3.88 (1824)	4.21 (1514)	4.12 (1315)	4.06 (1248)	3.86 (897)	3.84 (692)	3.72 (519)	3.66 (486)	3.50 (239)	4.11 (216)	3.28 (207)	3.34 (91)
Nonwhite(5)	3.09 (2388)	4.92 (2174)	4.71 (1709)	4.04 (1584)	4.26 (1112)	4.58 (1021)	4.33 (913)	4.03 (878)	3.64 (383)	3.75 (284)	3.58 (270)	3.34 (254)	2.68 (215)	4.18 (91)	3.94 (66)	3.32 (65)	4.87 (46)	4.17 (46)	2.78 (36)	1.33 (3)
Income: L. T. $15,000(6)	3.38 (1542)	5.97 (1360)	5.68 (1140)	5.07 (1021)	5.18 (795)	5.58 (689)	5.28 (537)	4.93 (514)	4.2[illegible] (298	5.11 (230)	5.18 (199)	5.56 (180)	5.06 (125)	4.68 (94)	4.85 (61)	4.81 (59)	2.53 (19)	2.95 (19)	2.40 (15)	6.00 (6)
$15–25,000(7)	3.21 (1569)	5.29 (1452)	4.34 (1232)	4.20 (1101)	4.30 (840)	4.60 (752)	4.29 (636)	3.80 (611)	3.41 (304)	3.92 (256)	3.89 (220)	3.75 (205)	3.45 (145)	3.71 (96)	2.59 (68)	2.58 (62)	2.80 (30)	2.86 (28)	4.00 (26)	1.60 (10)
G. T. $25,000(8)	4.15 (1002)	5.96 (941)	5.18 (863)	4.29 (770)	4.58 (639)	4.39 (582)	4.38 (511)	3.71 (496)	3.71 (279)	3.64 (231)	3.65 (204)	3.80 (202)	3.84 (147)	3.44 (129)	3.38 (103)	2.37 (98)	3.58 (38)	4.35 (34)	2.00 (34)	2.67 (12)
Education: High School(9)	3.70 (1073)	5.74 (1001)	5.12 (815)	4.82 (729)	5.19 (563)	5.39 (502)	4.71 (432)	4.44 (410)	3.75 (241)	4.16 (196)	4.34 (178)	4.94 (166)	3.41 (128)	3.66 (70)	3.22 (51)	3.84 (51)	2.88 (32)	3.22 (31)	4.00 (28)	3.64 (11)
Some College(10)	3.92 (1530)	6.19 (1385)	5.31 (1220)	5.03 (1079)	4.96 (825)	5.50 (727)	5.10 (605)	4.34 (583)	4.54 (259)	7.07 (207)	6.16 (187)	4.69 (173)	4.39 (124)	4.78 (102)	4.57 (70)	3.69 (64)	3.76 (33)	3.07 (30)	2.52 (27)	2.18 (11)
College(11)	3.03 (1734)	5.33 (1571)	4.58 (1383)	3.83 (1246)	4.17 (1013)	4.22 (911)	4.04 (749)	3.58 (724)	3.44 (421)	4.02 (352)	3.74 (292)	3.62 (280)	4.17 (189)	3.46 (164)	2.93 (120)	2.31 (111)	2.50 (24)	4.36 (22)	1.33 (21)	2.86 (7)
Age: 0–4 Years(12)	5.46 (978)	6.98 (862)	5.96 (727)	5.83 (650)	5.92 (493)	4.98 (418)	4.62 (334)	3.99 (316)	4.60 (181)	3.71 (150)	3.63 (118)	3.12 (105)	2.92 (89)	3.02 (53)	2.10 (40)	2.14 (28)	2.82 (17)	3.06 (17)	3.71 (14)	
5–19 Years(13)	2.58 (3056)	3.74 (2860)	3.79 (2473)	3.17 (2251)	3.36 (1803)	3.17 (1633)	3.65 (1442)	3.11 (1400)	3.47 (682)	3.10 (541)	3.38 (501)	2.98 (483)	2.67 (356)	3.57 (242)	3.66 (190)	2.99 (186)	2.79 (119)	3.70 (108)	2.59 (102)	1.85 (39)
20–44 Years(14)	3.43 (4613)	5.59 (4145)	5.18 (3531)	4.65 (3124)	4.94 (2384)	5.10 (2087)	4.82 (1757)	4.20 (1672)	3.77 (869)	4.09 (693)	3.95 (602)	3.98 (550)	4.14 (382)	3.89 (286)	3.75 (207)	3.85 (192)	4.27 (87)	4.15 (78)	2.89 (72)	4.48 (25)
45–54 Years(15)	3.82 (1208)	6.65 (1150)	5.19 (1026)	4.90 (896)	4.75 (724)	4.98 (668)	4.70 (584)	4.22 (565)	3.77 (301)	4.91 (255)	4.46 (217)	4.37 (217)	3.82 (176)	3.80 (120)	3.73 (104)	3.51 (98)	4.74 (54)	5.33 (54)	3.92 (50)	4.54 (22)
55 + Years(16)	4.20 (875)	6.41 (834)	5.65 (715)	4.88 (634)	5.25 (535)	5.88 (484)	5.40 (407)	4.73 (408)	4.75 (241)	6.28 (208)	5.56 (194)	5.93 (193)	4.74 (141)	4.65 (105)	4.68 (65)	4.85 (66)	3.83 (24)	3.62 (21)	5.52 (21)	3.43 (7)
Employer: I(17)	3.52 (1011)	6.42 (933)	6.08 (853)	5.25 (826)	5.27 (794)	5.18 (615)	5.42 (569)	4.40 (553)	3.87 (537)	4.22 (464)	4.04 (431)	4.27 (426)	4.32 (312)	4.14 (287)	3.78 (256)	3.65 (242)				
II(18)	5.41 (664)	6.48 (584)	4.75 (584)	4.77 (583)	4.42 (580)	5.11 (502)	4.61 (494)	4.02 (493)	3.67 (492)	4.37 (362)	3.51 (360)	3.32 (360)	3.14 (358)	3.49 (118)	2.96 (115)	2.50 (110)	2.94 (109)	4.22 (109)	2.89 (101)	
Cohort: A(19)	3.95 (269)	5.17 (269)	5.89 (269)	5.11 (269)	5.38 (269)	4.86 (269)	4.92 (269)	4.13 (269)	3.60 (269)	3.82 (269)	3.95 (269)	4.16 (269)	4.44 (269)	4.21 (269)	3.81 (269)	3.57 (269)				
B(20)	7.34 (200)	7.06 (200)	5.90 (200)	4.52 (200)	5.44 (200)	4.28 (200)	4.08 (200)	3.86 (200)	3.80 (200)	3.38 (200)	4.20 (200)	4.18 (200)	3.72 (200)	3.46 (200)						
C(21)	4.24 (338)	7.35 (338)	4.84 (338)	4.65 (338)	4.44 (338)	4.46 (338)	4.00 (338)	3.61 (338)	3.28 (338)	4.00 (338)	3.21 (338)	3.21 (338)	2.82 (338)							
No Survey	2.79 (6294)	4.60 (5574)	4.36 (4654)	3.84 (4078)	3.74 (3265)	3.70 (2737)	3.56 (2265)	3.39 (2130)	3.72 (1220)	3.68 (882)	3.43 (783)	2.94 (728)	2.81 (520)	3.43 (297)	3.21 (238)	3.36 (212)	2.14 (84)	1.23 (73)	2.65 (68)	1.06 (31)

We have chosen to analyze our data in a clear, straightforward manner so as to avoid some of the methodological problems of data adjustment that may have existed in past studies. First, we have not adjusted our data for seasonality or any other temporal factors. (A separate examination of the data indicated that seasonal influences are minor in comparison with the overall changes in utilization taking place in the start-up period.) Second, we present and analyze our results in simple graphic and tabular forms that can be easily understood and verified. Finally, we have adopted two conventions to increase the clarity of the results: 1) Our plots of utilization rates over time exclude data for population sample sizes below 100; and 2) First quarter utilization rates are excluded because they were depressed to meaningless levels by differential times of enrollment during the first quarter, difficulties of new enrollees in accessing the system, waiting times for initial appointments, etc. (Complete data histories and population sample sizes are given in Table 1.)

Results

Total Population Start-Up

Our results indicate the existence of a substantial start-up effect for the total study population. The characteristics of the start-up are illustrated in Figure 1, in which the average number of provider visits each quarter (annualized) are plotted against the length of plan enrollment in quarters. As can be seen, there is a clear downward trend in utilization. The duration of the start-up effect is notable; the downward trend appears to continue through quarter 19. The magnitude of the start-up is also appreciable. There were 5.37 annualized visits in quarter 2, but only 3.23 visits in quarter 19, yielding an absolute reduction in the level of utilization of over two visits during the 5-year study period. This represents a 40 per cent reduction in utilization.‡‡

Perusal of Figure 1 will indicate that the reduction in the utilization rate is relatively larger in the earlier quarters and tends to slow down after quarter 8 or 9. This reduction is also indicated in the first row of Table 2 (Study Population). Columns A, B and C show the average number of provider visits per quarter (annualized) for quarters 2, 9 and 13, respectively. Column D shows the *percentage* reduction in utilization that occurs between quarters 2 and 9, and column E indicates the percentage reduction in utilization between quarter 2 and quarter 13. The *absolute* reduction in the number of visits between quarters 2 and 9, and between quarters 2 and 13, are displayed in columns F and G, respectively.

Referring to these figures, we see that the percentage reduction between quarters 2 and 9 was 28.30 per cent, representing an absolute reduction of 1.52 visits. The comparison of quarters 2 and 13 yields only slightly larger numbers, 32.77 per cent, or 1.76 visits, indicating the slower rate of reduction that appears to continue to the end of the plot at quarter 19.

---

tions of this are not clear, however, since many other factors influence actual access importantly, such as administrative delays, variations in staff availability and appointment waiting time, and these factors vary over time and between plans. In our study, the quarter measures seem to yield virtually the same result; the start-up plot for the total population using a close approximation to the exact quarter definition is identical in all key respects to the plot in Figure 1. This may not be true in all cases, however, so we urge the consistent application of the calendar quarter measure in future studies.

‡‡ Straightforward tests of statistical significance cannot be performed on these differences in utilization rates because, owing to the definition of cohorts in the longitudinal analysis, the study population size decreases in each successive quarter. We have chosen not to adjust the population to facilitate testing because it would be confusing and would provide little relevant information that is not already contained in the graphical and tabular presentations.

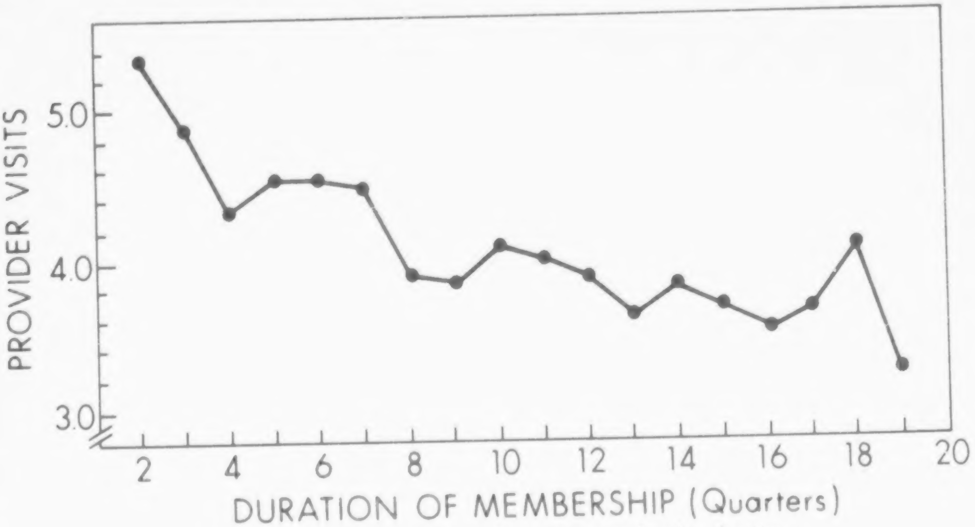


FIG. 1. Utilization rate of enrollee versus membership duration.

Close inspection of Figure 1 will also indicate several upward irregularities in the start-up trend. These can be seen as temporary increases in utilization that center on or about quarters 6, 10, 14 and 18. The periodicity of these "bumps"—occurring as they do every four quarters after quarter 2—suggests the existence of a systematic cause, a topic that will be discussed later in this paper.

Start-Ups Among Subgroups

Our results also indicate the existence of a pronounced start-up effect for different subgroups of the total population that are defined by sex, race, income, education and age, as well as for different employers and cohort groups. The characteristics of these different start-ups, however, vary between subgroups in several ways.

Sex. The utilization patterns for the female and male subgroups of our study are shown in Figure 2. Beginning with quarter 2, there is a definite downward trend in utilization that is once again interrupted by periodic bumps that peak during or close to quarters 6, 10, 14 and 18. The durations of the start-up effects, however, seem to differ between females and males. Although the bumps confuse the graphic interpretations somewhat, the females' rate of utilization appears to flatten out and vary at a roughly constant level after quarter 9. The start-up trend for males, on the other hand, appears to continue through quarter 16. The numbers given in row 2 of Table 2 (Female) also reflect the apparent development of this "steady state." The reductions between quarter 2 and quarter 9 of 1.54 visits, or 26.28 per cent, are nearly as large as those obtained four quarters later in the comparison between quarter 2 and quarter 13 (1.61 visits, or 27.47 per cent). The continuous reductions in male utilization are also suggested by the figures; the 30.5 per cent reduction in utilization achieved between quarters 2 and 9 is followed by a 39.0 per cent reduction between quarters 2 and 13.

Race. Figure 3 shows the start-up effects in the utilization behavior of the racial subgroups of "white" and "nonwhite" enrollees. Looking first at the start-up plot for whites, the rates of reduction appear to be

TABLE 2. Changes in Annualized Number of Provider Visits over Time

	(A) Q_2	(B) Q_8	(C) Q_{13}	(D) % Reduction in Utilization: Q_2 to Q_8	(E) % Reduction in Utilization: Q_2 to Q_{13}	(F) Absolute Reduction in Utilization: Q_2 to Q_8	(G) Absolute Reduction in Utilization: Q_2 to Q_{13}
Study Population(1)	5.37	3.85	3.61	28.30	32.77	1.52	1.76
Sex:							
Female(2)	5.86	4.32	4.25	26.28	27.47	1.54	1.61
Male(3)	4.82	3.35	2.94	30.50	39.00	1.47	1.88
Race:							
White(4)	5.52	3.88	3.86	29.71	30.07	1.64	1.66
Nonwhite(5)	4.92	3.64	2.68	26.02	45.53	1.28	2.24
Income:							
L. T. $15,000(6)	5.97	4.34	5.06	27.30	*	1.63	*
$15–$25,000(7)	5.29	3.41	3.45	35.54	*	1.88	*
G. T. $25,000(8)	5.96	3.71	3.84	37.75	*	2.25	*
Education:							
High School(9)	5.74	3.75	3.41	34.67	*	1.99	*
Some College(10)	6.19	4.54	4.39	26.66	*	1.65	*
College Degree(11)	5.33	3.44	4.17	35.46	*	1.89	*
Age:							
0–4 Years(12)	6.98	4.60	2.92	34.10	*	2.38	*
5–19 Years(13)	3.74	3.47	2.67	7.22	*	0.27	*
20–44 Years(14)	5.59	3.77	4.14	32.56	*	1.82	*
45–54 Years(15)	6.65	3.77	3.82	43.31	*	2.88	*
55 + Years(16)	6.41	4.75	4.74	25.90	*	1.66	*
Employer:							
I(17)	6.42	3.87	4.32	39.72	*	2.55	*
II(18)	6.48	3.67	3.14	43.36	*	2.81	*
Cohort:							
A(19)	5.17	3.60	4.44	30.37	*	1.57	*
B(20)	7.06	3.80	3.72	46.18	*	3.26	*
C(21)	7.35	3.28	2.82	55.37	*	4.07	*

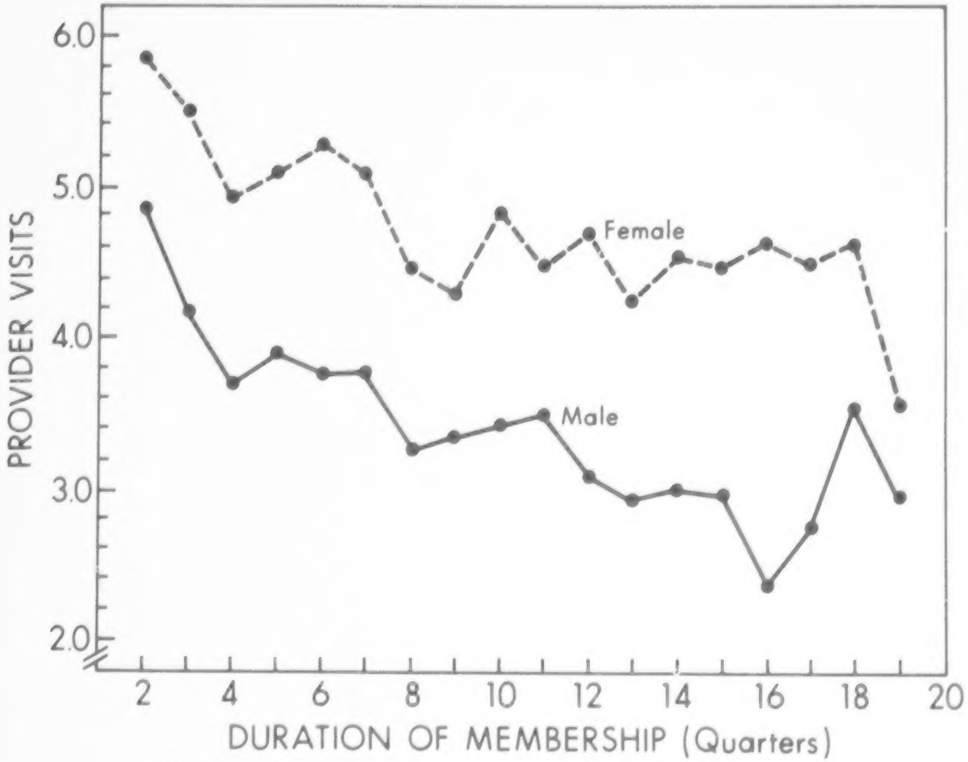


FIG. 2. Sex: utilization rate of enrollee versus membership duration.

greater in the earlier quarters, followed by a more gradual downward trend of long duration that persists through quarter 19. The slowing of the downward trend is also illustrated in Table 1; comparing the numbers for quarters 2, 9 and 13 shows that most of the reduction in provider visits occurred early. Once again, the bumps in utilization center in or around quarters 6, 10, 14 and 18.

The start-up pattern for nonwhites is generally more precipitous than that for whites. This can be seen both in Figure 3 and in Table 1. The overall reduction in number of visits between quarters 2 and 13 is substantially greater for nonwhites. Figure 3 also indicates that, for most of the history, the rate of utilization in absolute terms is greater for whites than for nonwhites. As can be seen, bumps also appear in the nonwhite graph at quarters 6 and 10.

Income. Figure 4 shows the start-up plots for three subgroups of subscribers defined by three different family income categories ("less than $15,000 per year," "$15,000 to $25,000," "and greater than $25,000"). Unfortunately, the greater variability in the data associated with smaller sample sizes reduces clarity and begins to confuse our graphic analyses in several ways. First, the trends themselves are harder to determine. Second, the bumps in utilization seem to peak at slightly different points, and the troughs no longer coincide. This also makes it hard to compare the magnitude of the utilization reductions between subgroups on the graphs; evaluating changes between quarters 2 and 9 re-

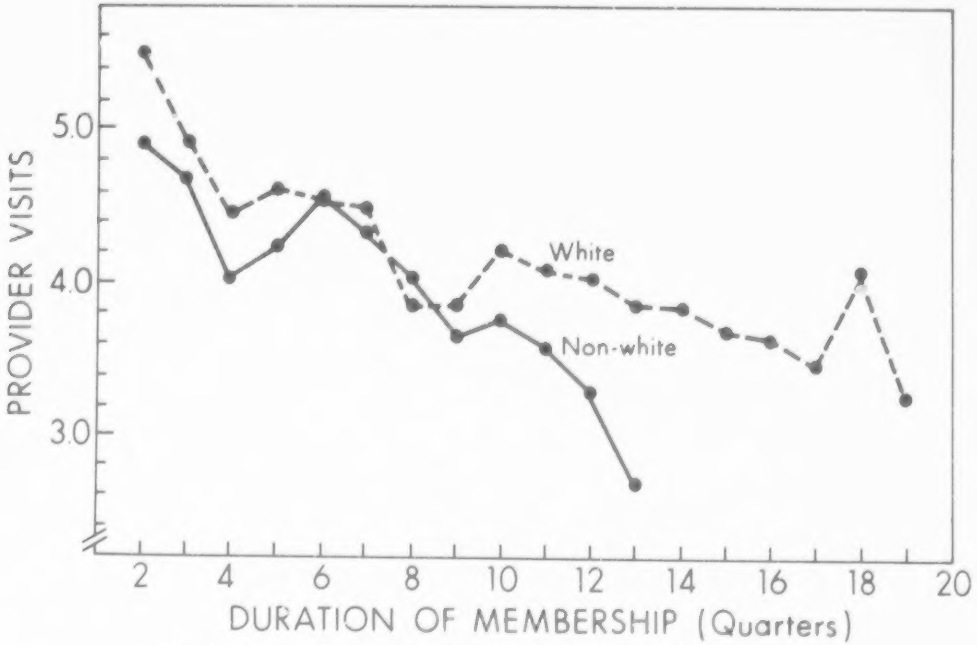


FIG. 3. Race: utilization rate of enrollee versus membership duration.

mains useful, but comparisons between quarters 2 and 13 are misleading.§§ Notwithstanding the increasing difficulty of interpretation, some similarities and differences are evident. All three graphs show definite start-up effects occurring between quarters 2 and 9. The magnitude of the start-up is smaller for enrollees earning less than $15,000 than it is for others. The duration of the start-up for those earning less than $15,000 is also different, apparently ending after quarter 9, whereas the other two start-ups seem to continue to quarter 16. Finally, the rates of utilization for these two higher-income categories are roughly the same over time, both being well below the utilization of individuals earning less than $15,000.

Education. The start-up plots for three different subgroups defined by the education level of the subscriber ("high school or less," "some college" and "college degree or more") are presented in Figure 5. As with income subgroups, all three plots seem to show a start-up effect from quarter 2 through quarters 8 or 9. Then the "some college" graph no longer decreases systematically. The other two start-ups seem to continue through quarters 15 and 16. As to the magnitude of the start-ups, the figures in Table 2 indicate that the subgroups of "high school or less" and "college degree or more" had roughly comparable results (34.67 per cent and 35.46 per cent, respectively) and that these were appreciably greater than those of the subgroup "some college" (26.66 per cent). In terms of absolute magnitude of utilization, "some college" and "high school or less" are comparable through quarter 8, while "college degree or more" shows consistently lower rates through quarter 12. Finally, note that the bumps peak at quarters 5 or 6 and quarters 10, 11 or 12.

---

§§ An asterisk (*) appears in columns E and G of Table 2 in these cases.

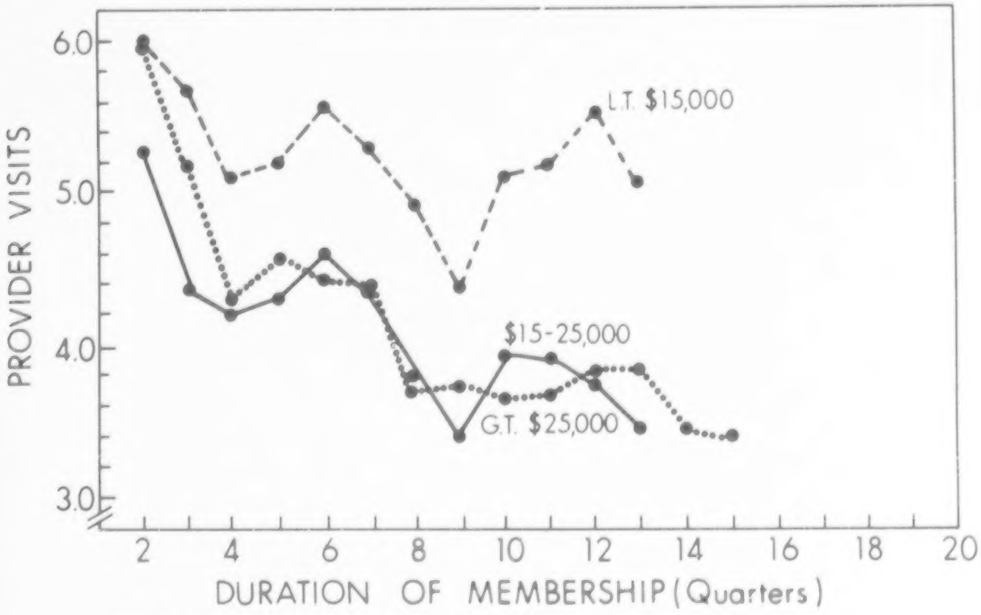


FIG. 4. Family Income: utilization rate of subscriber versus membership duration.

Age. Start-up plots are shown in Figure 6 for the following five age categories: 0 through 4 years of age, 5 through 19, 20 through 44, 45 through 54, and 55 or older. The age category of 0 through 4 years has the most dramatic start-up. It continues from quarter 2 to the end of the plot in quarter 12. The magnitude is appreciable, yielding reductions of 34.10 per cent between quarters 2 and 9. Rather well-defined bumps occur in quarters 5 and 9.

The age category of 45 through 54 years has the next most pronounced start-up effect, with a clear downward trend to quarter 9, by which time there has been a reduction of 2.88 annualized visits, or 43.31 per cent. Bumps occur at quarters 6 and 10.

The plot for the age category of 20 through 44 years shows a start-up effect through quarter 9, with a reduction between quarters 2 and 9 of 1.82 visits, or 32.56 per cent.

The data for the age category of 55 years and older document a very irregular utilization pattern; the number of annualized provider visits varies between roughly 6.5 and 3.5 from quarter to quarter, thus reflecting the highest level of utilization of any age group during much of the 5-year period. Although there is a hint of a start-up between quarters 2 and 9, the large bump that peaks at quarter 6 and the series of abruptly higher utilization levels that occur after quarter 9 confuse our graphic interpretation.

The plot for the age group of 5 through 19 years shows an attenuated start-up. As can be seen, the number of visits seems to drift downward irregularly from quarter 2 to quarter 13 before the plot becomes quite erratic. Note that for most of their history the absolute levels are well below those of other age groups.

Employer and Cohort Groups

For purposes of planning, PGP managers will be interested in the start-up ef-

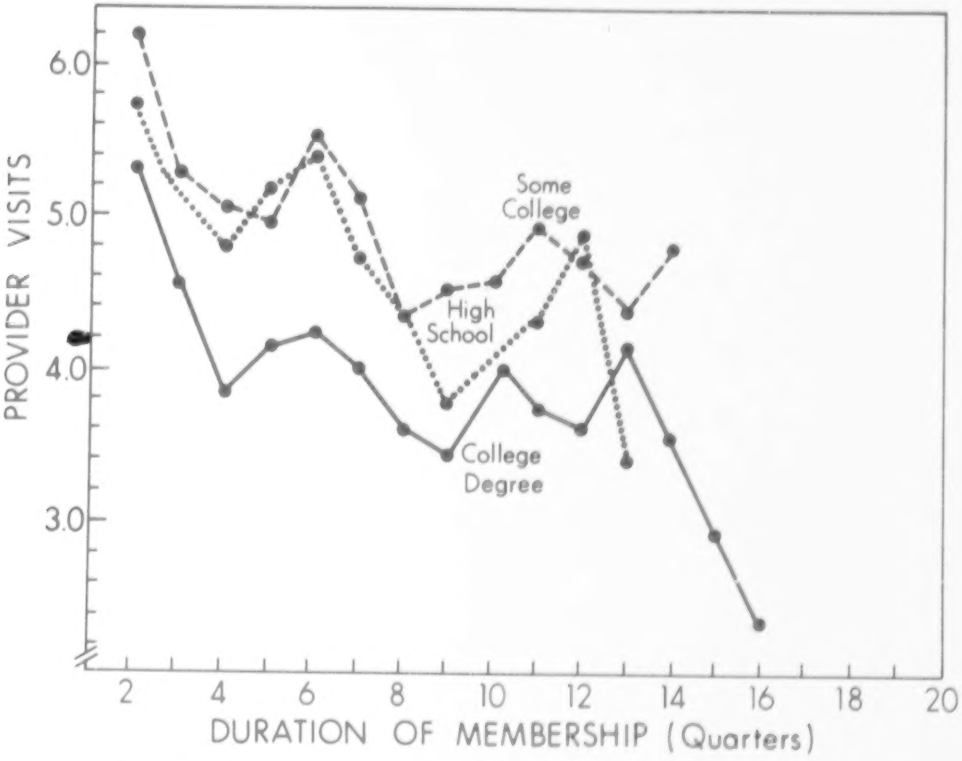


FIG. 5. Subscriber's Education: utilization rate of subscriber versus membership duration.

fects exhibited by natural enrollment groups that are drawn from the same employer or by cohorts of individuals who join the plan at the same time. We have, therefore, examined the start-ups of two employer groups that were among the largest and earliest to contract with MCG and those of three calendar cohorts who were among MCG's largest and earliest subscribers.

Figure 7 shows that the start-up effect exists for both employers. The downward trends in utilization are again disrupted by bumps that peak at quarters 6 or 7, 10 and 13 or 14, and 18. The duration of the start-up is appreciable in both graphs, with the employer I plot showing a downward trend throughout its entire history. The employer II start-up seems to continue through quarter 16. The rates of reduction in provider visits appear to be roughly the same early in the start-up, but after quarter 9 the employer I graph slows, while the graph for employer II continues downward rapidly. This is also seen in Table 1.

The start-up graphs for the three different cohorts are shown in Figure 8. The rates of decline for cohorts B and C seem to be somewhat greater than that for cohort A initially, but by quarter 9 the rates for all three groups are within 0.52 annualized provider visit of one another. The magnitude of the reductions between quarters 2 and 9 are all large, with visits being reduced by 1.57, 3.26 and 4.07, for percentage reductions of 30.37, 46.18 and 55.37 per cent, respectively, for A, B and C. Cohort A, however, appears actually to

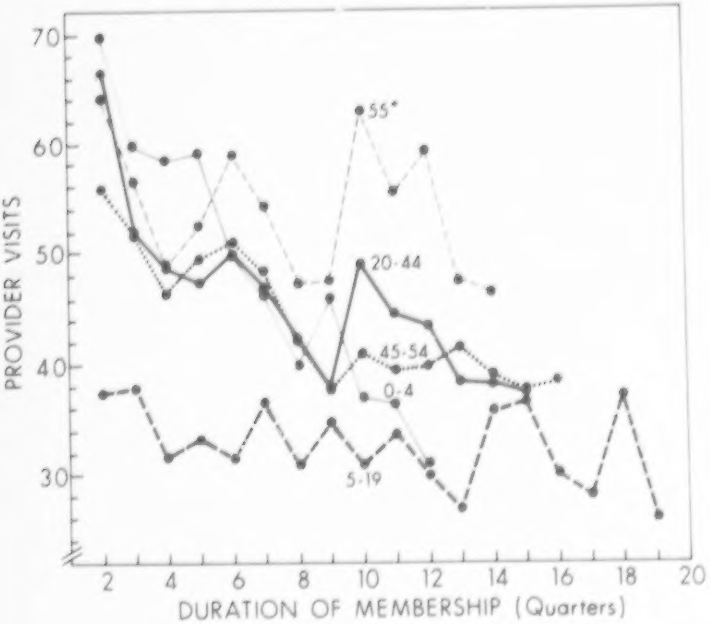


FIG. 6. Age: utilization rate of enrollee versus membership duration.

begin its start-up at quarter 3 instead of quarter 2, for reasons that we cannot explain. If, for this reason, we use the quarter 3 value and compare it with quarter 9, there is a reduction of 2.29 visits, or 38.88 per cent, which is closer to that achieved by B and C. Note also that bumps in utilization occur in each graph but that peaks are not located together, occurring as they do in quarters 5, 6, 7, and 10 or 11.

Discussion

The results reported above document the existence of a major start-up effect whose strength and persistence are notable; the study population as a whole and virtually all subgroups of the study population had appreciable reductions in utilization rates over the 5-year study period, and these results also held across different cohorts and employers.

How general or widespread is the phenomenon? Although this is an empirical question, some factors argue for generality. Consider first its causes. We expect that the behaviors underlying the start-up phenomenon are normal consequences of the utilization incentives inherent in prepaid group practice. With reduced access barriers, enrollees can cause high initial utilization by testing the benefits of the plan, by having periodic checkups for which they are now covered, and so on.[2] The PGP organization can also stimulate high initial utilization by encouraging initial screenings, detecting and treating preexisting illnesses, encouraging greater awareness of health status and providing health education to enrollees, and through other administrative practices and policies.

The comparability of our study PGP with many others in the United States also suggests a widespread start-up effect. For instance, the Medical Care Group has roughly 24,000 enrollees, making it comparable in size to over 100 other PGPs in the United States. It is a relatively young organization, having been in operation for 10 years. Approximately 200 other plans

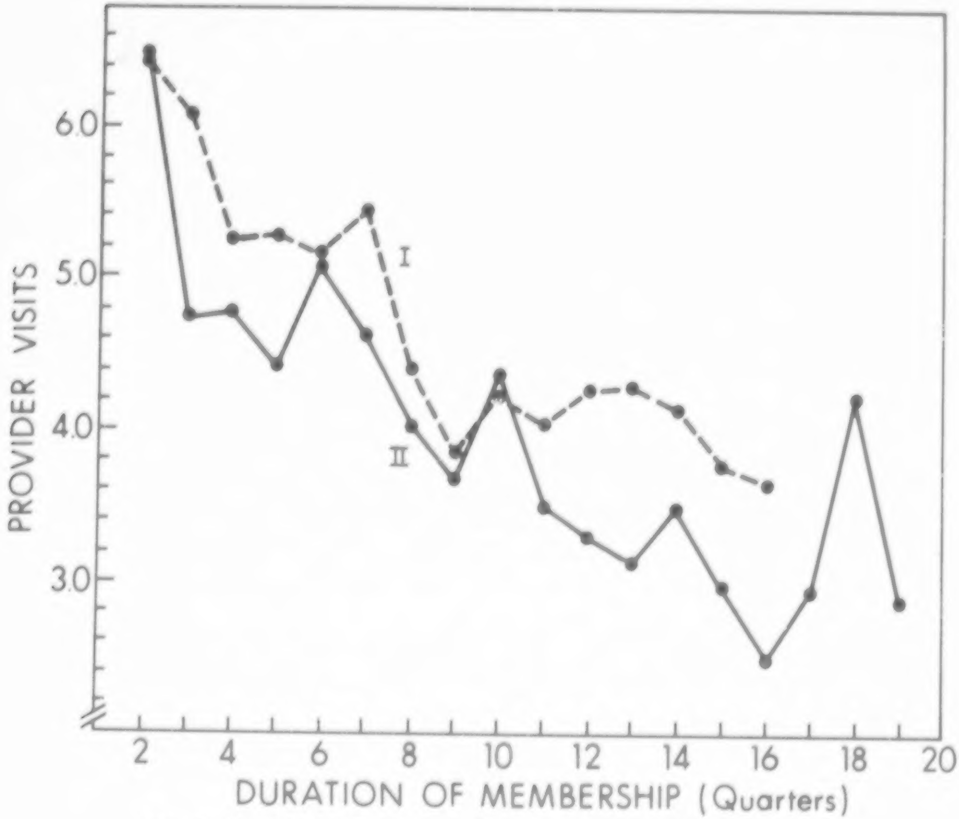


FIG. 7. Employer: utilization rate of enrollee versus membership duration.

have been in operation less than 10 years, and 50 of these are between six and nine years old. Overall utilization figures are also similar. MCG members, taken as a whole, have roughly 3.6 provider visits per year. This is quite comparable to the overall average of 3.4 visits by other PGPs in 1979.[12] MCG is also similar to other plans in its broad SES composition, urban location, broad benefit package and absence of financial mechanisms that could restrict utilization.

We do not presume, however, that sizable start-ups will occur in all PGPs. Just as plans are able to stimulate high initial utilization, they can also dampen it rather effectively through the use of different policies or administrative practices. These practices could include applying such financial barriers as deductibles, restricting the range of covered benefits, discouraging initial physical examinations, allowing long waiting lines to develop and reducing access in other ways. To the degree that such practices exist, they might be more prevalent in older and larger PGPs that have had more time to implement them and less financial concern about positive incentives for enrollment growth.

In summary, then, we suggest as a *working hypothesis* to guide future studies that start-ups will be found widely, especially

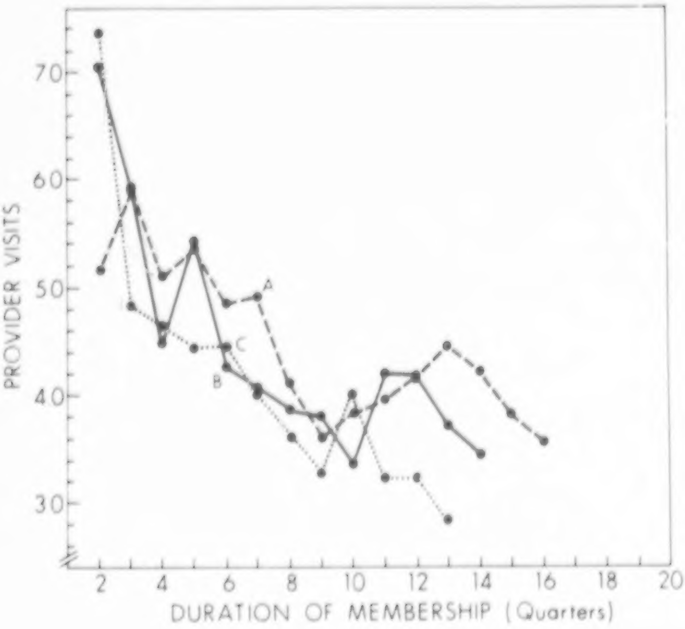


FIG. 8. Cohort: utilization rate of enrollee versus membership duration.

among young and developing PGPs. Their magnitude, duration and other characteristics may, however, vary appreciably from plan to plan.

The periodic bumps in utilization found throughout our results also deserve discussion. Recall that these bumps occurred roughly every four quarters after quarter 2—centering around quarters 6, 10, 14 and 18—and that they account for a sizable amount of utilization.

We were not able to analyze directly or document the causes of this phenomenon. We have some strong suspicions, however, that lead us to suggest, as another working hypothesis, that they are caused by the federally mandated procedure of notifying employees annually of their health benefits and open enrollment dates. We suspect that these annual notifications could be stimulating increased utilization in one or more of the following ways: 1) some persons are reminded to follow up on previous treatment; 2) others may be induced to take further advantage of their prepaid benefits through, for example, scheduling a physical examination or consulting a physician about their general health status; 3) the notification may heighten the anxiety of some "worried well" and thereby stimulate a temporary increase in their utilization of health care services. Other examples could be given, but the hypothetical mechanism would be the same: annual notification could increase enrollee awareness of their medical benefits and their health status, and this heightened awareness could be translated into more provider visits. More visits could result in the treatment of yet undiscovered illness with follow-up visits. The overall result of such a process could cause the bumps in utilization that occur in our data roughly every 12 months.

Implications

One of the more important policy implications of the start-up effect concerns financial incentives for enrolling new members. The rate of diffusion of prepaid group practice would perhaps be increased if some form of reimbursement were provided for the costs of a start-up. Assume for argument's sake that future research shows that significant start-up effects normally occur in most PGPs and that associated costs are large. Administrators might then be reluctant to expand their plans beyond financially viable or break-even levels because enrolling new members would either increase the average per capita cost, implying higher premiums, or reduce the real service capacity (per enrollee) of the plan. In either case there is a disincentive for plan expansion. Furthermore, such disincentives may be selectively larger for some socioeconomic groups than for others. This type of pernicious effect could result, for instance, if there is a belief or expectation that certain racial minorities or socioeconomic groups have higher levels of unmet medical needs that would translate into larger or longer start-ups.

One possible solution to this problem would be to provide a financial incentive for the enrollment and start-up of new groups that goes above the average premium. If done with private funds, for example, PGPs could be urged to charge a "start-up premium" to new employers during the first years of enrollment. With public funds it would be necessary to develop some start-up incentive to be applied above the established rates of reimbursement by Medicare, Medicaid or other forms of public financing.

We will not be able to judge the appropriateness or effectiveness of such start-up incentives until a number of important questions are answered by future studies. For example, does the higher initial rate of utilization during a start-up represent effective and appropriate care, or is it actually *overutilization* that should be discouraged? If it is appropriate, how can the size and duration of the incentives be established in an equitable manner? Other conceptual problems also exist, though most of them are similar to those usually encountered in public financing of health care delivery.

The regular bumps that occurred in our start-up data could also have policy implications. If future study confirms our hypothesis that open enrollment notification and the annual statement of benefits are responsible for such bumps, it would be wise to reconsider the merit and form of this requirement. Once again, there are some nettlesome questions: do the bumps represent effective and appropriate utilization, or only some form of overutilization related to anxiety? In either case, the *form* of the current procedure needs to be rethought if it is resulting in unintended health utilization behavior. If the bumps represent inappropriate utilization, ways should be found to minimize or eliminate these behavioral consequences. If the utilization is appropriate, perhaps the method of notification should be redesigned to strengthen its motivational effect.

In addition to policy implications, start-ups have important administrative consequences for a large variety of activities related to the design, planning and control of PGPs. In conducting these activities, planners and managers rely heavily on some average rates of utilization of services by their enrolled population, such as the number of visits per enrollee per year. This average utilization rate becomes, in managerial parlance, a "standard" for translating expected patient demand for services into the required physical facilities, medical staff and equipment necessary to provide these services. Hence, staffing decisions, such as the number of physicians, physician's assistants, RNs, technicians and administrative personnel to employ at any point in time, are made using these standards. They are also used in a large number of other planning and monitoring activities. Rate setting is one obvious example, since these standards of utilization are employed to estimate the base from which capitation rates and premiums are ultimately determined. Marketing activities are also affected because the standards are used to estimate how much unused capacity or delivery capability a plan would be able to reserve for new enrollees in the future. These are just a few of the activities that rely on standards of utilization.

References

1. Broida JH, et al. Impact of membership in an enrolled, prepaid population on utilization of health services in a group practice. N Engl J Med 1975;292:780.

2. Yesalis CE, Bonnet PD. The effect of duration of membership in a prepaid group health plan on the utilization of services. Med Care 1976;14:1024.

3. Mullooly JP, Freeborn DK. The effect of length of membership upon the utilization of ambulatory care services: a comparison of dis-advantaged and general membership populations in a prepaid group practice. Med Care 1979;17:922.

4. Avnet HH. Physician service patterns and illness rates. Group Health insurance, Inc., 1967.

5. Robertson RL. Comparative medical care use under prepaid group and free choice plans: a case study. Inquiry 1972;9:70.

6. Shragg H, et al. Low-income families in a large scale prepaid group practice. Inquiry 1973;10:52.

7. Sparer G, Anderson A. Utilization and cost experience of low-income families in four prepaid group practice plans. N Engl J Med 1973;289:67.

8. Forthofer RN, Glasser JH. Utilization of services of an HMO by new enrollees. AJPH 1979;69:1127.

9. Baloff N, Becker SW. On the futility of aggregating individual learning curves. In: Schultz, ed. The Science of Psychology: Critical Reflections. New York: Appleton-Century-Crofts, 1970.

10. Perkoff GT. Changing health care: perspectives from a new medical care setting. Ann Arbor; Health Administration Press, 1979.

11. Wilcox V. Summary: 1978 Medical Care Group survey. Unpublished monograph. Division of Health Care Research, Washington University, Spring 1980.

12. National HMO Census of Prepaid Plans: 1979. U.S. Department of Health, Education and Welfare, Public Health Service, Office of Health Maintenence Organizations, Division of Program Promotion.

MEDICAL CARE
December 1981, Vol. XIX, No. 12

Communications

Commentary on "Membership Duration and Utilization Rates in a Prepaid Group Practice," by Mary Jane Griffith and Nicholas Baloff

JOHN P. MULLOOLY, PH.D.*

GRIFFITH AND BALOFF present important data on the use of health services by new members of a young, average-sized, urban PGP program. However, some problems with their methodology raise questions about the validity of their conclusions.

The question of the magnitude of error introduced by using the approximate calendar quarter definition of duration of membership rather than the exact definition employed by Mullooly and Freeborn and by Forthofer, Glasser and Light is probably of minor importance compared with other issues raised by the methodology of this article. Before discussing these other matters, it is worth mentioning that if one applies the approximate calendar quarter definition of duration of membership, the actual number of person–months of membership should be used in the denominator of the first-quarter rate. Since there is no information on the relative frequency of monthly enrollments during the first quarter of membership, the adjusted first-quarter rates, assuming a uniform distribution (i.e., an average membership of two months), become 5.22 for respondents, 4.19 for nonrespondents and 4.84 overall. Computing first-quarter rates this way contributes to a more complete description of the start-up phenomenon and allows for comparisons of actual first-quarter rates among PGPs.

The following methodological problems should be dealt with in designing future studies of start-up effects in PGPs: 1) selection biases due to excluding disenrolled members and nonrespondents from the analyses; 2) inherent statistical variability of utilization rates; and 3) variations in the seasonal mix of the study population by duration of membership due to uneven quarterly enrollments. These problems have not been addressed by the authors of this study.

In order to facilitate discussion of these methodological problems, I have reconstructed the enrollment pattern of the study population. Table A shows the numbers of members who joined MCG during the calendar quarters of 1973 through 1977. These data were obtained by successive subtractions of the respondent and nonrespondent population figures given in Table 2 (in Griffith and Baloff). For example, the 98 respondents who had 20 quarters of membership joined during the first quarter of 1973. The difference between

---

* Senior Investigator, Health Services Research Center, Kaiser Foundation Hospitals, Portland, Oregon.

0025-7079/81/1200/1221/$00.75 © J. B. Lippincott Co.

TABLE A.

Calendar Quarter	No. Enrollees				Calendar Quarter	No. Enrollees				Calendar Quarter	No. Enrollees			
	Total	Resp.	Non-resp.	% Non-resp.		Total	Resp.	Non-resp.	% Non-resp.		Total	Resp.	Non-resp.	% Non-resp.
1973 1	129	98	31	24.0	1975 1	622	404	218	35.0	1977 1	2,429	1,616	813	33.5
2	198	161	37	18.7	2	129	84	45	34.9	2	1,493	917	576	38.6
3	24	19	5	20.8	3	314	215	99	31.5	3	2,299	1,379	920	40.0
4	34	23	11	32.4	4	765	427	338	44.2	4	1,599	879	720	45.0
										Total	17,024	10,730	6,294	37.0
1974 1	397	269	128	32.2	1976 1	2,997	2,087	910	30.4					
2	62	36	26	41.9	2	298	163	135	45.3					
3	259	200	59	22.8	3	1,238	766	472	38.1					
4	561	338	223	39.8	4	1,177	649	528	44.9					

the number of respondents with at least 19 quarters of membership and those with 20 quarters (i.e., 259 − 98 = 161) represents the number who joined during the second quarter of 1973.

By moving these enrollment cohorts through time, the correspondence between calendar quarter or season and quarter of membership can be determined. For example, the 20th quarter of membership for the 98 respondents who joined during JFM of 1973 occurred during OND of 1977. Following these 98 respondents back through time, it is seen that their 19th quarter of membership occurred during JAS of 1977, the 18th during AMJ of 1977, the 17th during JFM of 1977, the 16th during OND of 1976, etc. Table B shows the respondent and total study populations by duration of membership and calendar quarter of season.

It is seen from Table A that the percentage of nonrespondents increased from approximately 20 per cent for the first three enrollment cohorts of 1973 to approximately 40 per cent for the last three enrollment cohorts of 1977. From Table 2 (in Griffith and Baloff), it may be seen that the percentage of nonrespondents decreases with duration of membership. Nonrespondents constituted approximately 35 per cent of the population with 1–6 quarters of membership and approximately 22 per cent of the population with 17–20 quarters of membership. Given this association between response rate and duration of membership, and the fact that nonrespondents had lower rates than respondents, analysis of the relationship between utilization rate and duration of membership should be based on the entire study population, not just on the respondents.

Since the authors have not provided information on the variability of the rates, I have estimated the standard deviation of an annualized utilization rate to be approximately six visits per year, based on comparable data reported by Forthofer and

TABLE B.

Duration of Membership (Quarter)	Calendar Quarter (Season) 1 (J, F, M) No.	100 p1	2 (A, M, J) No.	100 p2	3 (J, A, S) No.	100 p3	4 (O, N, D) No.	100 p4	Total	%
1 R	4,474	41.7	1,361	12.7	2,579	24.0	2,316	21.6	10,730	100.0
R + NR	6,574	38.6	2,180	12.8	4,134	24.3	4,136	24.3	17,024	100.0
2 R	1,437	14.6	4,474	45.4	1,361	13.8	2,579	26.2	9,851	100.0
R + NR	2,537	16.4	6,574	42.6	2,180	14.1	4,134	26.8	15,424	100.0
3 R	1,200	14.2	1,437	17.0	4,474	52.8	1,361	16.1	8,472	100.0
R + NR	1,835	14.0	2,537	19.3	6,574	50.1	2,180	16.6	13,126	100.0
4 R	444	5.9	1,200	15.9	1,437	19.0	4,474	59.2	7,555	100.0
R + NR	687	5.9	1,835	15.8	2,537	21.8	6,574	56.5	11,633	100.0
5 R	2,858	48.1	444	7.5	1,200	20.2	1,437	24.2	5,939	100.0
R + NR	4,145	45.0	687	7.5	1,835	19.9	2,537	27.6	9,204	100.0
6 R	788	14.9	2,858	54.0	444	8.4	1,200	22.7	5,290	100.0
R + NR	1,360	16.9	4,145	51.6	687	8.6	1,835	22.9	8,027	100.0
7 R	434	9.6	788	17.4	2,858	63.2	444	9.8	4,524	100.0
R + NR	597	8.8	1,360	20.0	4,145	61.1	687	10.1	6,789	100.0
8 R	281	6.4	434	10.0	788	18.1	2,858	65.5	4,361	100.0
R + NR	389	6.0	597	9.2	1,360	21.0	4,145	63.9	6,491	100.0
9 R	771	33.9	281	12.4	434	19.1	788	34.7	2,274	100.0
R + NR	1,148	67.1	389	11.1	597	17.1	1,360	38.9	3,494	100.0
10 R	361	19.5	771	41.7	281	15.2	434	23.5	1,847	100.0
R + NR	595	21.8	1,148	42.1	452	16.6	597	21.9	2,729	100.0
11 R	219	13.4	361	22.1	771	47.2	281	17.2	1,632	100.0
R + NR	283	11.7	595	24.6	1,148	47.5	389	16.1	2,415	100.0
12 R	197	12.7	219	14.1	361	23.3	771	49.8	1,548	100.0
R + NR	260	11.4	283	12.4	595	26.0	1,148	50.2	2,286	100.0
13 R	367	32.1	197	17.2	219	19.1	361	31.6	1,144	100.0
R + NR	526	31.6	260	15.6	283	17.0	595	35.8	1,664	100.0
14 R	23	2.9	367	45.5	197	24.4	219	27.2	806	100.0
R + NR	34	3.1	526	47.7	260	23.6	283	25.7	1,103	100.0
15 R	19	3.1	23	3.8	367	60.6	197	32.5	606	100.0
R + NR	24	2.8	34	4.0	526	62.3	260	30.8	844	100.0
16 R	161	28.2	19	3.3	23	4.0	367	64.4	570	100.0
R + NR	198	25.3	24	3.1	34	4.3	526	67.3	782	100.0
17 R	98	32.6	161	53.5	19	6.3	23	7.6	301	100.0
R + NR	129	33.5	198	51.4	24	6.2	34	8.8	385	100.0
18 R	0	0.0	98	35.3	161	57.9	19	6.8	278	100.0
R + NR	0	0.0	129	36.8	198	56.4	24	6.8	351	100.0
19 R	0	0.0	0	0.0	98	37.8	161	62.2	259	100.0
R + NR	0	0.0	0	0.0	129	39.4	198	60.6	327	100.0
20 R	0	0.0	0	0.0	0	0.0	98	100.0	98	100.0
R + NR	0	0.0	0	0.0	0	0.0	129	100.0	129	100.0

R = respondents.

R + NR = respondents + nonrespondents.

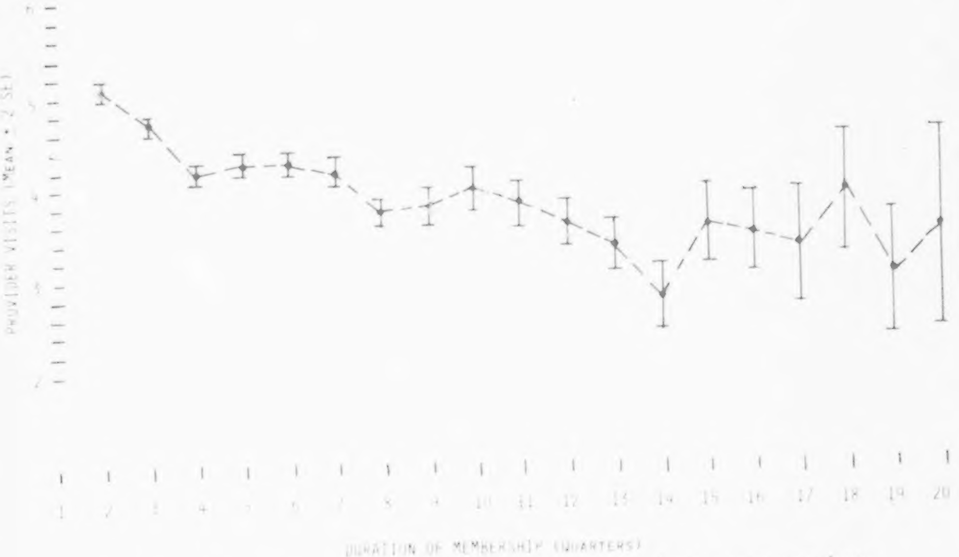


FIG. A. Utilization rate versus membership duration; respondents + nonrespondents.

Glasser. Using this estimate, I have placed approximate 95 per cent confidence limits around the mean rates for the study population (respondents and nonrespondents) (Fig. A). It is seen that the apparent decline in provider visit rate is limited to the first eight quarters of membership and that the magnitude of the variability inherent in the

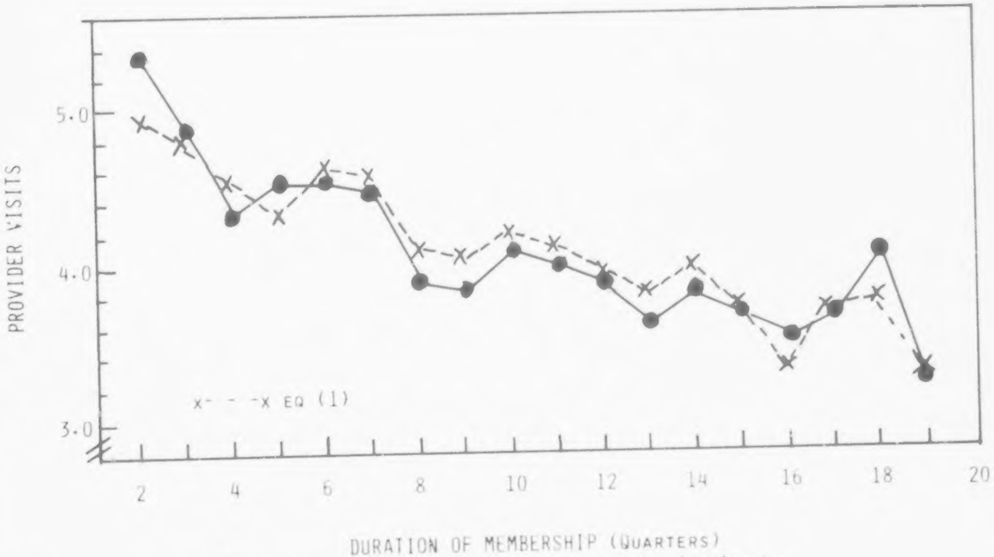


FIG. B. Utilization rate of respondents versus membership duration.

utilization rates makes it inappropriate to attach importance to subsequent drops in the rate.

Furthermore, the decrease over the early quarters of membership may in part be due to a selection bias introduced by excluding members who terminated or disenrolled during the study period. Although MCG is reported to have had a low disenrollment rate, from 2 to 5 per cent of the population per year, it should be noted that the majority of disenrollments from PGP programs occur among members with short durations of membership and that the probability of disenrollment decreases with time in the program. Also, unpublished data from the Kaiser Foundation Health Plan of Oregon indicate that terminated members with short durations of membership have substantially lower utilization rates than members who remain in the program. The expected effect of including disenrolled members in the analysis would be a moderation of the decline in utilization rate over the early quarters of membership. Valid assessment of the magnitude of the start-up phenomenon requires elimination of this selection bias.

Paired tests based on within-subject comparisons of level of use during intervals of duration of membership provide an appropriate means of assessing the statistical significance of changes in utilization rates.

The authors suggest that the observed pattern of periodic "bumps" in their utilization data was caused by annual notification of health benefits and open enrollment dates. An alternative explanation supported by their data is that this periodic variation is attributable to seasonal variation in utilization. From Table B it is seen that there was a great deal of variation in the seasonal mix of the MCG population by duration of membership and that this variation was periodic, with a period of four quarters of membership. To illustrate this point quantitatively, the sequence of utilization rates for respondents during quarters 2 through 20 was fit by ordinary least squares to a linear function of duration of enrollment (qrt) and the proportions (p1, p2, p3, p4) of the respondent population for which the qrtth quarter of membership occurred during JFM, AMJ, JAS, OND, respectively. The resulting regression equation,

$$\text{rate} = 5.143 - .081\ \text{qrt} - 0.598\ \text{p1} + 0.360\ \text{p2} - 0.562\ \text{p4} \quad \text{(Eq. 1)},$$

accounts for 87.8 per cent of the variation in the utilization rates. Figure B compares the plots of the regression estimates with the observed rates shown in Figure 1 (in Griffith and Baloff). It is seen that the variation in the seasonal composition of the population produced a pattern of bumps which closely fits the observed pattern.

The large variation in the number of quarterly enrollments during the period of 1973–1977 seen in Table A produced the large variation in the seasonal mix of the population by duration of membership. PGPs that enroll approximately equal numbers of members each calendar quarter would be expected to have approximately uniform seasonal distributions except for large durations within a year of the maximum duration defined by the length of the study period.

Even small seasonal differences in utilization, which are acknowledged by the authors, can produce periodic variation in the utilization rates by duration of membership when there is wide variation in the seasonal mix of the population by duration of membership.

Discussion of national policy and HMO administrative implications of the relationship between membership duration and utilization should be based on sound examination of the underlying phenomenon.

MEDICAL CARE
December 1981, Vol. XIX, No. 12

Commentary on "Membership Duration and Utilization Rates in a Prepaid Group Practice," by Mary Jane Griffith and Nicholas Baloff

CHARLES E. YESALIS, III, M.P.H., SC.D.,* AND SUSAN D. HORN, PH.D.†

GRIFFITH AND BALOFF's study of the effect of duration of membership on the use of physician services‡ in a prepaid group practice (PGP) further supports the hypothesis that per capita utilization declines from its initial high level and stabilizes thereafter. Their study examines the effects of income, education and race, variables that have been ignored by previous research in the area. Furthermore, the generalizability of the research appears to be substantial, in that the Medical Care Group (MCG) mirrors many of the characteristics of PGPs in this country.

These important findings notwithstanding, Griffith and Baloff's study does have two major difficulties. The first is that the authors have assumed that the MCG provided a "stable" research environment during the time of the study. The second is the adequacy of the time trend analysis within subpopulations.

Research Environment

Organizations are not static—changes take place over time, and, in our opinion, PGPs and their membership are no exception. Many of these changes directly affect the use of medical services and, thus, could potentially confound the study of the effect of duration of membership. Some of the more important potential correlates of utilization that might change over time are:

1. Need for medical services;
2. Health education levels of enrollees;
3. Provider availability and accessibility;
4. Use of ancillary personnel;
5. Provider mix;
6. Amount and types of benefits available to plan members;
7. Training, competence and experience of providers; and
8. Use of deductibles, copayments and coinsurance.[1-3]

It is difficult to imagine that the MCG or any PGP did not experience some significant fluctuation in the above variables over a 5-year period. However, Griffith and Baloff have assumed exactly that and in

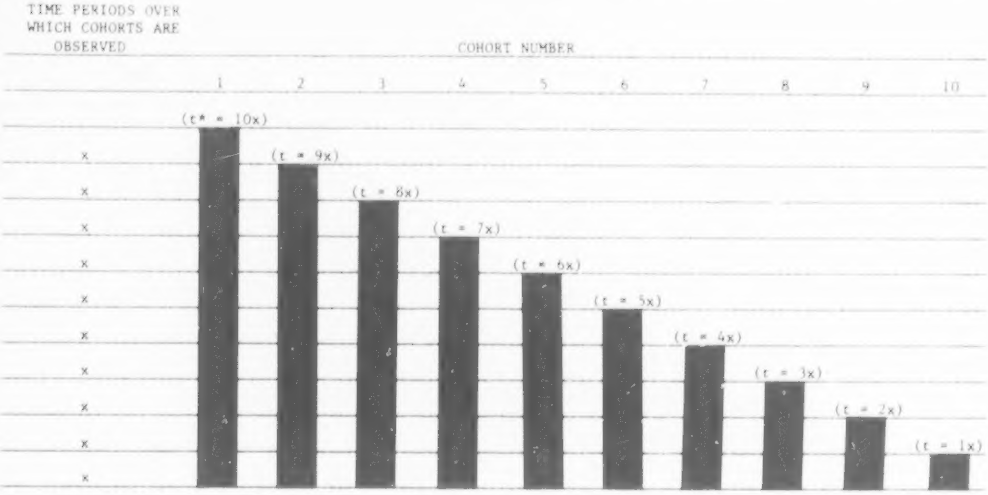
---

* Associate Professor, College of Medicine, University of Iowa, Iowa City, Iowa.

† Associate Professor, School of Hygiene & Public Health, Johns Hopkins University, Baltimore, Maryland.

‡ This also includes visits to paraprofessionals.

0025-7079/81/1200/1226/$00.85 © J. B. Lippincott Co.



*t = total time period over which cohort is observed

FIG. 1.

effect have attributed most of the changes in utilization time trends to duration of membership. One exception is the periodic "bumps" in utilization observed by the authors approximately every fourth quarter. Such systematic fluctuations in observations over time are consistent with what time trend analysts consider seasonal fluctuations. The authors state, however, that "seasonal influences are minor . . . (and) we have not adjusted our data for seasonality or any other temporal factors." This is confusing, considering the attention that the authors have shown these bumps, and considering the availability of accepted techniques (to be discussed below) to adjust for seasonal variations in time trends. Economists have employed such techniques for years when analyzing inflation and employment data. What is of even greater concern to this author is the potential number of nonsystematic changes in the MCG whose effects on utilization were not easily identifiable through Griffith and Baloff's graphic analysis. Moreover, any such shifts in utilization trends were likely explained as the effects of duration of membership.

In a study of the effect of duration of membership at the Columbia Medical Plan (CMP), a research method that was effective in identifying and controlling potentially confounding variables was identified.[4] This method involved multiple longitudinal cohort analysis (Fig. 1) and multiple dummy variable regression. In this illustration 10 cohorts were designated, and each was observed for a different length of time. Of course, any number of cohorts can be used, but it is desirable that the number of time periods be at least as great as the total number of cohorts.

The design in Figure 1 provides 10 observations on utilization of services during the first year of contact with the group, nine observations on the second year of contact, and so on, to one observation on the 10th year of contact (for cohort 1).

This multiple-cohort design permits one cohort to serve as a comparison for other cohorts, and it also helps to identify and differentiate any effects on utilization of

TABLE 1. PVs/Member/Quarter Year (Annualized Quarterly Rates) by Duration of Membership and by Cohort

Duration of Membership (quarters)	Cohort									
	1	2	3	4	5	6	7	8	9	10
1	4.70	5.53	7.48	4.91	5.97	4.36	4.30	5.08	5.33	4.86
2	5.36	5.35	6.95	7.23	8.27	6.44	6.23	5.69	7.00	6.32
3	5.31	5.10	6.63	7.33	5.75	5.22	7.57	6.72	6.52	5.28
4	5.17	4.95	6.70	6.45	5.55	5.08	7.20	6.56	5.60	6.89
5	5.59	5.08	6.14	5.20	5.49	5.34	6.66	4.80	7.33	
6	5.35	5.01	5.59	6.18	6.28	5.58	5.10	6.56		
7	4.98	4.67	6.51	6.39	7.32	5.12	6.39			
8	4.90	4.86	7.11	6.19	5.00	6.18				
9	4.92	5.55	6.62	4.58	6.47					
10	5.26	5.41	5.26	5.79						
11	5.32	4.72	5.22							
12	4.79	5.54								
13	5.62									

services associated with a specific study quarter (e.g., summer or winter) from those associated with duration of enrollment in the group. On the other hand, if a single-cohort design is employed as in Griffith and Baloff's study, events associated with chronological time of enrollment are held synonymous with the effects of duration of membership.

There are a number of changes that took place over time in CMP that might have confounded the effects of duration of membership on the use of services. Those changes that were of special concern in our prepaid group practice study were provider/enrollee ratio, provider mix, available office hours and need for medical care.

Seasonal fluctuation in need for care is used as an example here. In Table I we list the provider visits (PVs) per member per quarter for each cohort by duration of

TABLE 2. PVs/Member/Quarter Year (Annualized Quarterly Rates) by Study Quarter and by Cohort

Study Quarter	Cohort									
	1	2	3	4	5	6	7	8	9	10
1	4.70									
2	5.36	5.53								
3	5.31	5.35	7.48							
4*	5.17	5.10	6.95	4.91						
5	5.59	4.95	6.63	7.23	5.97					
6	5.35	5.08	6.70	7.33	8.27	4.36				
7	4.98	5.01	6.14	6.45	5.75	6.44	4.30			
8*	4.90	4.67	5.59	5.20	5.55	5.22	6.23	5.08		
9	4.92	4.86	6.51	6.18	5.49	5.08	7.57	5.69	5.33	
10	5.26	5.55	7.11	6.39	6.28	5.34	7.20	6.72	7.00	4.86
11	5.32	5.41	6.62	6.19	7.32	5.58	6.66	6.56	6.52	6.32
12*	4.79	4.72	5.26	4.58	5.00	5.12	5.10	4.80	5.60	5.28
13	5.62	5.54	5.22	5.79	6.47	6.18	6.39	6.56	7.33	6.89

* Quarter year period covering the months of July, August and September.

membership. No seasonal effects are obvious. However, a rearrangement of the data according to *study quarter* (as opposed to duration of membership) produces different results (Table 2). The data for study quarters 4, 8 and 12 all of which included the months July, August and September, show generally lower utilization rates than either the preceding or subsequent study quarters. This can probably be attributed to less need or demand for care, especially care related to upper respiratory illness, as well as to the number of providers and enrollees out of the area on summer vacation. As stated previously, Griffith and Baloff's data resemble the seasonal patterns presented here, but there is no way of knowing whether these configurations are seasonal effects or are explained by duration of enrollment.

Another example is shown in data on walk-ins§ per 100 members per quarter year. In Table 3 there is no apparent correlation between duration of membership and walk-ins, with the exception of the last two quarters of membership in each cohort. In these quarters a dramatic increase in walk-ins is noted. The probable cause of this increase is made clearer by displaying the rate of walk-ins by study quarter (see Table 4). It now appears that this increase is not associated with duration of membership, but rather that it is associated with the last two quarter years of the study.

At the start of the 12th study quarter, the CMP opened its own hospital, which was the first for the city in which the plan is located. Until this time, plan members had to drive at least 15 miles to reach a hospital emergency room. With the opening of the new emergency room, medical care was made more accessible, resulting in an increase in emergency room visits, the majority of which were classified as walk-ins. It appears that it was the opening of the hospital and not duration of membership that resulted in the large increase in walk-ins during study quarters 12 and 13.

The mere identification of confounding factors is not always sufficient, however. Thus, we will discuss next dummy variable regression, a procedure for adjusting utilization data to control confounding variables.

---

§ PVs that were not arranged in advance in any way.

TABLE 3. Walk-Ins/100 Members/Quarter Year (Annualized Quarterly Rates) by Duration of Membership in the CMP and by Cohort

Duration of Membership in the Plan (quarters)	Cohort									
	1	2	3	4	5	6	7	8	9	10
1	18.28	33.25	44.52	29.29	29.07	18.76	29.80	42.58	25.88	34.04
2	25.55	28.10	34.18	26.71	30.62	30.70	47.02	32.03	40.51	39.39
3	29.96	28.60	27.85	23.90	21.05	32.78	55.63	23.49	53.16	88.34
4	31.72	22.11	30.38	24.60	39.23	30.29	27.81	33.45	71.73	74.95
5	18.50	21.28	30.38	35.15	23.92	27.38	37.75	63.34	83.54	
6	22.91	27.10	36.71	34.45	23.92	39.83	65.56	55.52		
7	20.26	26.60	39.24	32.34	31.58	68.05	70.86			
8	29.07	26.77	30.38	34.45	56.46	64.73				
9	32.60	30.92	45.57	59.75	66.03					
10	28.19	27.43	78.48	74.52						
11	27.31	56.69	56.96							
12	56.39	62.18								
13	60.79									

TABLE 4. Walk-Ins/100 Members/Quarter Year (Annualized Quarterly Rates) by Study Quarter and Cohort

Study Quarter	Cohort 1	2	3	4	5	6	7	8	9	10
1	18.28									
2	25.55	33.25								
3	29.96	28.10	44.54							
4	31.72	28.60	34.18	29.29						
5	18.50	22.11	27.85	26.71	29.07					
6	22.91	21.28	30.38	23.90	30.62	18.76				
7	20.26	27.10	30.38	24.60	21.05	30.70	29.80			
8	29.02	26.60	36.71	35.15	39.23	32.78	47.02	42.58		
9	32.60	26.77	39.24	34.45	23.92	30.29	55.63	32.03	25.88	
10	28.19	30.92	30.38	32.34	23.92	27.38	27.81	23.49	40.51	34.04
11	27.31	27.43	45.57	34.45	31.58	39.83	37.75	33.45	53.16	39.39
12	56.39	56.69	78.48	59.75	56.46	68.05	65.56	63.34	71.73	88.34
13	60.79	62.18	56.96	74.52	66.03	64.73	70.86	55.52	83.54	74.94

Dummy variable regression is a simple and useful method that uses binary independent variables to incorporate categorical information such as season, study quarter and sex into a regression equation. It can also be used to adjust data for the effects of such variables. Although the technique is not new, it is usually not discussed in elementary text books on statistical methods. In our research the 13 quarter years of the study were entered into the regression equation as study quarters. A regression coefficient was then determined for each study quarter. This coefficient estimated the impact of that study quarter on the dependent variable. This technique has been proposed before by Price,[5] but the study-quarter effect is usually called the period effect and the duration-of-membership effect is usually

TABLE 5. Regression Coefficients for Provider Visits Per Member Per Quarter Year by Study Quarter

Study Quarter	Coefficient	F	Proportion of Sample	Adjusted Coefficient (Coefficient − Q)
C	5.054			4.251
a_1	-2.752	15.811	0.013	-1.949
a_2	-1.925	7.955	0.013	-1.122
a_3	-1.872	13.155	0.026	-1.069
a_4	-1.378	9.676	0.039	-0.575
a_5	-0.935	5.492	0.053	-0.132
a_6	-0.405	1.214	0.066	+0.398
a_7	-1.073	9.724	0.079	-0.270
a_8	-1.447	19.831	0.092	-0.644
a_9	-0.913	8.723	0.105	-0.110
a_{10}	-0.301	1.032	0.118	+0.502
a_{11}	-0.284	0.992	0.132	+0.519
a_{12}	-1.342	22.959	0.132	-0.539
a_{13}	0		0.132	+0.803

Q = .803

called the age effect. The data on PVs/member/quarter year in Table 1 are observations on the dependent variable Y; the independent variables are dummy variables representing the study quarters S_i, i = 1, . . . , 13. We omitted S_{13} from the equation and estimated the coefficients c and a_i, i = 1, . . . , 12 for the regression equation.

$$Y = c + a_1 S_1 + \ldots + a_{12} S_{12} + e,$$

where as usual e is a random error component. The coefficients and their F values are given in Table 5. The adjusted coefficients are found by solving the equation

$$\sum_{i=1}^{13} P_i \, (a_i + Q) = 0$$

for Q; the a_i are the original coefficients, the P_i are the corresponding proportions in the sample and Q is the adjusting constant used to calculate the adjusted coefficients incorporating the effect of S_{13}. This is the procedure suggested by Sweeney and Ulveling.[6] The adjusted coefficients are then used to adjust the PV data for the effects of confounding variables that occur in the study quarters. The adjusted PV utilization data are then analyzed to determine the relationship between utilization of services and duration of membership free of the spurious effects of confounding variables in the study quarters.

Regarding this method, one could question the use of a cohort analysis when the adjustments for confounding effects on utilization of services could be dealt with by the dummy variable regression alone. It is true that the adjustments to the dependent variable (i.e., use of services) could be made without the multiple-cohort analysis, but the researcher would have difficulty in determining the purpose for which the adjustments were being made. It should be noted that use of the "cohort" component of the design does not guarantee identification of all confounding variables. However, with sufficient supplementary data, such as manpower and total number of enrollees, a number of important influences on utilization of services can be brought to light. Furthermore, the researcher should develop, wherever possible, the theoretical constructs behind significant statistical associations and should not just adjust data "in the dark."

Time Trend Analysis

In their study Griffith and Baloff spend considerable time describing differences in time trends within subpopulations (e.g., the utilization patterns of males vs. females over time), especially differences in the duration of the "start-up" effect. These differences appear often, in our opinion, to be in the eye of the beholder or, in cases where a concensus of viewers may be achieved, the differences could be attributed solely to random variation. Since no tests of statistical significance are performed, the reader is left to fend for himself or herself. There are statistical models and corresponding computer programs available, however, that may be appropriate for these data.[4-7] We used the exponential decay curve model in our study at the CMP, and it appears that this model would apply to Griffith and Baloff's data as well. The equation for the exponential decay curve is:

$$y = c + ae^{-bx} + \epsilon,$$

where y = dependent variable; c = constant to be estimated (this represents the level to which the curve is asymptotic); a = constant to be estimated (y = c + a + e when x = 0); e = base of the natural system of logarithms (2.718); b = decay constant; and ϵ = residual error.

There are no closed-form solutions for the estimator of the parameters c, a and b as there are for the parameters in a linear regression equation. A computer program is used to obtain the least squares estimates of the parameters by an iterative technique, i.e., values of c, a and b were found that minimized $\Sigma[y_i - (c + ae^{bx} i)]^2$.

This model enables the researcher to differentiate statistically curvilinear time trends among subgroups. In addition, the model allows the health administrator or planner to estimate more accurately the point of stabilization of utilization (i.e., the asymptote) than could be achieved by the visual inspection method offered by Griffith and Baloff.

References

1. Aday L, Eichhorn R. Utilization of health services: indices and correlates. National Center for Health Services Research, DHEW, 1973.

2. Donabedian A. Benefits in medical care programs. Cambridge: Harvard University Press, 1976.

3. Donabedian A. Aspects of medical care administration. Cambridge: Harvard University Press, 1973.

4. Yesalis C, Bonnet P. The effect of duration of membership in a prepaid group health plan on the utilization of services. Med Care 1976;14:1024.

5. Price D. A respecification of variables in cohort analysis. Presented at the Annual Meeting of the American Statistical Association, Boston, Mass., August 1976.

6. Sweeney R, Ulveling E. A transformation for simplifying the interpretation of coefficients of binary variables in regression analysis. American Statistician 1972;26:30.

7. Horn S, Yesalis C, Bartels R. Selection of regression models for health care data. Med Care 1978; 16:574.

Medical Care
December 1981, Vol. XIX, No. 12

Reply to Comments: "Membership Duration and Utilization Rates in a Prepaid Group Practice"

Nicholas Baloff, Ph.D., and Mary Jane Griffith, Ph.D.

Our reaction to the two commentaries and to the guest editorial is decidedly mixed. On one hand, we are pleased by the extra attention that the start-up effect is receiving. On the other hand, we are disappointed by the focus and content of these remarks, for several reasons: 1) Our primary findings about start-ups and their implications have been largely ignored; 2) The authors of the commentaries have failed to respond to important criticisms of their own papers that were published earlier in this journal[1,2]; 3) Instead of dealing with either of these problems, the commentaries dwell upon methodologic criticisms of secondary issues; and 4) As a result, one receives the mistaken impression that our paper was the first in a sequence (rather than the third) and that the role of the commentaries was to critique our methodology.

A review of the past sequence of events will hopefully set the matter straight. In a paper published in this journal in 1976,[1] Yesalis and Bonnet reported evidence of a substantial start-up effect. In a second paper, published in this journal in 1979,[2] Mullooly and Freeborn disputed that finding, considering the start-up effect merely an artifact of the methodologic adjustments made by Yesalis and Bonnet. In addition, based on their own analysis of different data, Mullooly and Freeborn reported finding *no* start-up effect. Our review and reanalysis of the Mullooly and Freeborn paper indicated that their general conclusion seemed to be at odds with their own data, since four separate age–sex subgroups of the population appeared to have experienced a substantial start-up similiar to those found in our study. This apparent contradiction seems to have resulted from their methodologic decision to average the utilization data of different age–sex subgroups that had quite different utilization trends over time (some decreasing, others increasing) to yield an "average curve" for the entire population that may not have described any of the subgroups of which it was composed.

These contradictory findings and the questionable methodologic techniques upon which they were based greatly influenced our choice of objectives and methods: specifically, it seemed appropriate for us to return to the basic questions of the existence of the start-up phenomenon, its elementary characteristics and its generality across different subgroups of our study population. To avoid compromising the findings through analytic and methodologic manipulations, we chose the simplest, most descriptive forms of analysis, disclosed our data fully and avoided the types of data adjustments that had confused past results. With this background, the specific criticisms contained in the commentaries can be evaluated in proper context.

Mullooly Commentary

In his commentary, Mullooly argues several different points, including calendar quarter methodology, "selection bias" related to nonrespondents and disenrollees, duration of the start-up and seasonal-

0025-7079/81/1200/1233/$00.85 © J. B. Lippincott Co.

ity (as a potential cause of the periodic bumps in the start-up data). Judging from the space devoted to them, his two main concerns are start-up duration and the bumps. We shall therefore consider these first.

With respect to start-up duration, Mullooly reconstitutes and reanalyzes our data for the start-up of the entire study population and asserts that: "It is seen that the apparent decline in provider rate is limited to the first eight quarters of membership and that the magnitude of the variability inherent in the utilization rates makes it inappropriate to attach importance to subsequent drops in rate." This assertion is incorrect on two counts. First, it is *not* supported by the evidence he presents, as we shall show below. But more importantly, he has completely ignored the compelling contrary evidence that is contained in the subgroup start-ups. This is the same error that he made in his earlier paper.[2] He has focused narrowly upon the total population and overlooked the results in the subgroup analyses. He therefore draws an erroneous conclusion that is contradicted by the subgroup results. To illustrate the point, perusal of the start-ups for males and females in Figure 2 of our paper will show that the start-up for males *continues rather clearly through quarter 16*. The start-up for females, on the other hand, seems to end after quarter 8. Combine the two, and the resulting trend naturally "flattens" and becomes harder to interpret. But is it sensible to argue, as does Mullooly, that the start-up has *ended* after quarter 8—given that the males' start-up clearly continues well beyond that? Similar evidence can be found in the start-ups for racial subgroups (Fig. 3). It appears that both of these start-ups also continue beyond quarter 8, though it is clearer for nonwhites than for whites. Much additional evidence of start-up duration well beyond quarter 8 is contained in the other subgroup analyses. The key point, however, is the importance of looking beyond the aggregated, total population curve to the subgroup start-ups in drawing conclusions about the duration (or other characteristics) of the start-up phenomenon. It is also essential that we analyze subgroup start-ups to gain a better understanding of the phenomenon and its causes. It is, for example, important to acknowledge the different durations of start-ups between males and females if we wish to explore the causal mechanisms underlying them. It is also important for planners and administrators interested in practical applications to consider the different durations and other characteristics of start-ups between sexes (and races, etc.) in order to make appropriate estimates of start-ups for different populations—composed of different subgroups in varying proportions.

Even if, for the purposes of argument, one grants Mullooly his narrow focus on the total population curve, the evidence he presents fails to support his assertion. Based upon an assumed standard deviation (which is incorrectly held constant over time) and a reconstituted population of respondents and nonrespondents, he develops a figure showing mean utilization rates with ever increasing confidence limits. That is all! He does not statistically test this result in any way. He simply asserts in a completely judgmental manner that the decline in utilization rates is limited to the first eight quarters. This is clearly not valid. The variability of the utilization rates *naturally* increases with decreases in sample size, making it harder to evaluate an underlying downward trend. The obvious disruptions of the bumps also contribute to this difficulty. But this certainly does not *prove* that the trend ends in quarter eight. In fact, one can readily marshall statistical evidence to the contrary, in the following way: To reduce some of the variability after quarter 8, we aggregated the quarterly data into yearly periods. Mean rates for the full year following quarter 8 (quarters 9–12) and for the last year of the data (quarters 17–20) for both respondents and nonrespondents were found to be 4.570 and 4.148, with standard errors of 0.0678 and 0.1886, respectively. To determine whether these means and standard errors are consistent with continued reductions after quarter 8, we treated the observations in the last four quarters (17–20) and in the four quarters following quarter 8 (9–12) as independent (as a first approximation) and calculated the difference between the means for these periods, and then divided this by the estimated standard error of the difference, obtaining a value of 2.1074. This is statistically significant at the 0.05 level. Based upon this test, it appears that the reductions in utilization rates did indeed continue beyond quarter 8.

To summarize, Mullooly's assertion is not supported by his own argument. It is clearly contradicted by the subgroup results that he has ignored and by our statistical test results.

As his second point, Mullooly seeks to demonstrate that the periodic bumps in our start-ups may have been caused by seasonal variation in utilization. He alleges that "An alternative explanation supported by their data is that this periodic variation is attributable to seasonal variation in utilization." This allegation is incorrect. We shall again demonstrate this both in reference to important flaws in his argument and through direct refutation with our own results.

Mullooly's lengthy analysis may appear impressive on first reading. Closer examination, however, shows it to be essentially self-evident and inconclusive. His regression results are the problem. Mullooly claims an impressively large coefficient of determination (r^2) and then uses his regression equation to produce a new start-up plot that parallels our original graph, including the bumps. On this result alone he rests his case. The difficulty is that the result is really a *tautology:* the large coefficient of determination was *virtually guaranteed* because of the number and type of independent variables he used, and it therefore demonstrates little or nothing by itself. By then taking this self-evident result and calculating values for a new plot with bumps, he has done nothing more than give us a graphic illustration of the tautology. Speaking technically, the p_i variables in his equation are cyclic, with periods of four quarters and four different phases. It is obvious that any cyclic wave form of period 4, like the bumps in the start-ups, would be very well fitted with any three of these variables and an intercept. By adding his duration variables (qrt), the downward trend is also well fitted. Hence, an impressively large r^2 is almost certain to result.

It is puzzling that Mullooly has neglected to tell us more about his regression results. For example, in addition to the toal r^2, we are very interested in the *partial* explanations offered by the different independent variables—especially the main variable of *duration* of membership. Could it be that duration explained most of the variation, and the seasonality variables very little? If so, it would of course argue *against* Mullooly's allegation. It could also contradict his earlier assertion that the duration of the start-up was only eight quarters.

We turn now to our own evidence in order to refute Mullooly's allegation. Remember that we clearly stated in our paper that seasonal influences were minor in the start-ups. The analysis presented in Figure A shows why we made that statement and should erase any lingering doubts on the subject. The utilization rates shown there have been weighted in such a way as to control for (or cancel) any seasonality effect.* Hence, if seasonality were the cause,

---

* We separated our population into four groups of individuals by the calendar quarter in which they

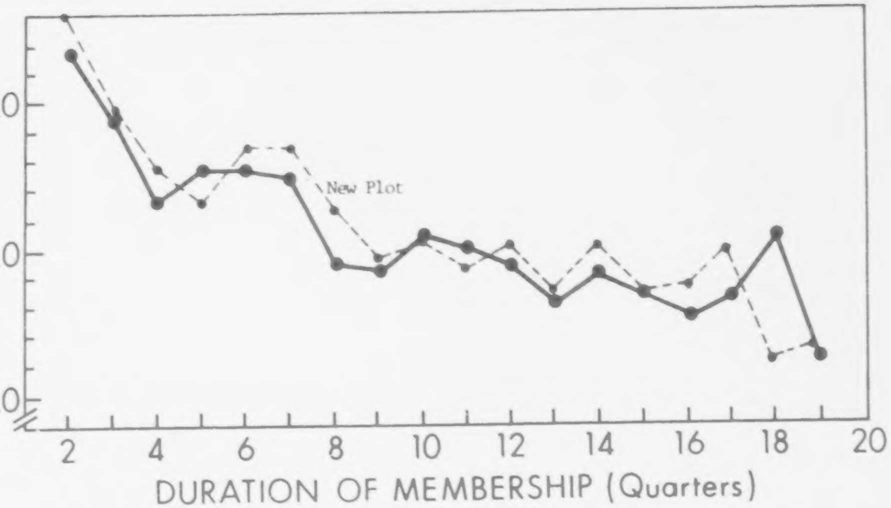


FIG. A. Utilization rate of enrollee versus membership duration.

no bumps would have appeared in the figure. As can be clearly seen, the bumps not only remain, they are still sizable, and their distribution over time is approximately the same as before.

To summarize, the second of Mullooly's two major points is not supported by his own results—and it is clearly refuted by our analysis.

We turn now to the remaining points in his commentary. Specifically, his comments about "calendar quarter" methodology—which he employed in his paper[2]—are *not* responsive to the criticisms of this approach contained in our paper. His detailed comments about "person-months of membership" are irrelevant to our primary criticism of his approach, as well as being incorrect (assuming an average membership of two months to adjust first-quarter rates is wrong). Next, the comments on "selection bias" fail to recognize key points made in our paper, as follows: 1) Nonrespondents were excluded from our analyses because SES data were not available for them. Including them would therefore have made it impossible to evaluate the start-ups for the different subgroups—which could lead to the types of erroneous conclusions that Mullooly has made in his paper and commentary. 2) However, as stated in our paper, we analyzed the start-up curve for the nonrespondents *separately* and found that it closely parallels that of the respondents. 3) Regarding disenrollment, Mullooly's concerns should be placed in the context of the very low 2–5% per year disenrollment rate that existed over the five-year study period. It seems unreasonable to claim that excluding such a small number of persons could have materially changed the start-ups—particularly given their large magnitude and consistency across subgroups.

Yesalis and Horn Commentary

In the second commentary, Yesalis and Horn argue three points: 1) We should have adjusted for seasonality and for vari-

---

enrolled (winter, spring, etc.), calculated mean utilization rates for each group for each quarter of membership, then averaged these to give equal representation to each group in the calculation of the seasonally adjusted mean utilization rates shown in the figure. This has the effect of cancelling seasonal effects by adjusting for different numbers of persons who enrolled in different seasons.

ous changes in the plan (provider-to-enrollee ratio, provider mix, available office hours, etc.); 2) We should have used their methodology of so-called "multiple cohorts" to compose the population for analysis rather than the so-called "single cohorts"; and 3) We should have employed statistical models to analyze our data and to determine the duration of the start-up effect. We shall discuss these three points in turn.

There are several reasons for rejecting the first point. The clearest and most compelling of these reasons is that the adjustments they are advocating so strongly are precisely those used by Yesalis and Bonnet in their 1976 paper and validly criticized by Mullooly and Freeborn[2] *for artificially producing a start-up trend* where none was evident in the raw utilization data for several of the cohort groups. This criticism is a powerful one, for it makes it seem that the start-ups reported by Yesalis and Bonnet were simply methodologic *artifacts* of their "adjustments"—the very same adjustments that Yesalis and Horn are now recommending to us. Furthermore, neither Yesalis and Bonnet (in their original article) nor Yesalis and Horn (in their commentary) have really examined this criticism, or explained how their "adjustments" could have produced such curious and troubling results. They have also not provided enough explication of these procedures to enable readers to reconstruct the details of the adjustment process and determine for themselves which adjustments created the problems, and how.

We are thus left with considerable uncertainty over both the validity of the original results and the credibility of the adjustment methods used to arrive at them. It seems completely inappropriate, therefore, to urge application of these same adjustment techniques to our study. On the contrary, the problems associated with them are partially responsible for our insistence on using simpler analyses to establish the credibility of the start-up effect and its "unadjusted" characteristics.

The second reason for rejecting their suggestion is the absence of a comprehensive theoretic basis for making such adjustments. Although it may *seem* that the variety of the plan changes cited by Yesalis and Horn could affect utilization behavior and rates, such presumptions are not adequate grounds for making multiple, interrelated adjustments to data—especially when the adjustments change the raw data so appreciably.

We turn now to their second issue relating to "multiple-cohort" versus "single-cohort" designs. Since Yesalis and Bonnet used the multiple-cohort design *exclusively* in their earlier paper,[1] it is not surprising that now Yesalis and Horn extol its benefits. What is very surprising—even baffling—is that they criticize us for using the single cohort instead of their multiple-cohort approach. It appears either that they have not read our paper completely, or that they did not comprehend it fully. We have specifically included *the very same* multiple-cohort approach they recommend, in addition to the single-cohort method. Figure 8 in our paper shows the start-ups for three cohorts of individuals that are defined *identically* to their "multiple cohorts." Furthermore, Figure 7 shows two additional start-ups of cohorts defined by employer group; these approximate the "multiple-cohort" definition closely because most of the individuals in these cohorts joined at the same time. It is unfortunate that Yesalis and Horn have overlooked these findings, because they show directly that their concerns about our analyses are groundless. Persual of Figures 7 and 8 will reveal that these start-ups have the same basic characteristics as seen in all the other plots using the "single-cohort" design. All five document large changes in utilization over long durations. Furthermore, the basic trends seem to resemble closely the other start-ups, and the evidence of underlying bumps is reasonably clear in most of them. These results provide considerable assurance that our "single-cohort" start-ups were not, as Yesalis and Horn criticized, affected by major changes in the plan or seasonal influences. In addition, although it is not as obvious, the stability of the single-cohort start-up characteristics across so many different recompositions of our population by SES subgroups also lends assurance on this point.

Given this evidence, Yesalis and Horn's criticism seems quite inappropriate and reflects a lack of comprehension of our results that is rather troubling under the circumstances. Furthermore, in their strong commitment to their own method, they have neglected to recognize the methodologic and practical benefits of the other method—and hence fail to realize the importance of employing *both* methods simultaneously as we have done in our study.

As their third point, Yesalis and Horn suggest the use of statistical models to analyze our data. They note the utility of regression analysis as a means of statistical testing, then remark on its usefulness in allowing administrators to estimate more accurately the point of stabilization of utilization at the end of a start-up. As we stated in our paper and reiterated at the beginning of this reply, both of these suggestions are basically inconsistent with the stated objectives and methodologic approach of the *present* paper. However, given the results of the present paper and the resulting clarification of the basic issues, regression analysis and start-up modeling are reasonable next steps. We have performed such analyses, and interested readers can find these results and their practical applications to many administrative functions discussed elsewhere.†

We began this reply by noting that both of the commentaries had ignored our major findings and the important criticisms of their own previous work, and instead had focused on narrow methodologic criticisms of secondary issues. We have since shown that their specific criticisms are largely incorrect or unsupported. We have also noted how both commentators seem to have revisited their own past errors—returning to the methodologies and analytic approaches that created problems in their original papers. This is unfortunate. We would all have been better served by a thoughtful consideration of the criticisms and a reanalysis of their past results and conclusions. This could have yielded a better understanding of the generality of start-ups across settings, how their characteristics may be modified by specific institutional factors and a number of other questions suggested in our paper.

Perrin Editorial

Perrin's editorial seems misdirected. He has applauded the two commentaries for being part of a process of "exploratory data analysis," but failed to recognize the serious errors that destroy their utility for this purpose. His philosophical remarks are therefore poorly illustrated by the commentaries. Indeed, these comments are much better illustrations (when combined with the authors' earlier articles) of the unfortunate consequences of applying statistical techniques and data adjustments without a sound understanding of the underlying phenomena.

References

1. Yesalis CE, Bonnet PD. The effect of duration of memberships in a prepaid group health plan on the utilization of services. Med Care 1976;14:1024.

2. Mullooly JP, Freeborn DK. The effect of length of membership upon the utilization of ambulatory care services: A comparison of disadvantaged and general membership populations in a prepaid group practice. Med Care 1979;17:922.

---

† See: Baloff N, Griffith MJ. Managing Startup Utilization in Ambulatory Care. Journal of Ambulatory Care Management (forthcoming). February 1982.

MEDICAL CARE
December 1981, Vol. XIX, No. 12

An Organized Locum Tenens Service and the Cost of Free Time for Rural Physicians

ALAN K. KRONHAUS, M.D.

Dissatisfaction with the conditions of rural practice perpetuates the geographic maldistribution of physicians. This study evaluates the economic productivity of physicians who participated in an organized locum tenens (LT) service that was designed to improve practice conditions for rural physicians by making time off more feasible and less costly. The productivity of LT physicians, expressed as their average daily billing, was compared with the amount charged to rural clinics for the LT service and with the revenue billed by the clinic physicians during the same month. The clinics were charged $375/regular working day for the LT service. LT physicians billed an average of $439/day, compared with $584/day billed by the clinic physicians ($p < 0.01$). The revenue billed by 85 per cent of the LTs exceeded the amount charged for their services; the revenue billed during 80 per cent of the coverage episodes also exceeded the $375/day level. When LT billing was expressed as a percentage of clinic physician billing, LT productivity was found to be significantly higher in solo practices, as compared with partnership practices (97 per cent vs. 73 per cent, $p < 0.01$). Assuming that a physician's overhead remains constant when he leaves, the cost of leaving a practice with LT coverage was almost invariably less than the cost of leaving without coverage. This finding justifies cautious optimism about using LTs to increase free time and improve practice conditions for rural physicians.

GEOGRAPHIC DISPARITIES in the per capita distribution of physicians exist in nearly every country. These disparities exist in part because most physicians prefer to live and work in metropolitan communities.[1,2] Moreover, dissatisfaction with the conditions of rural practice helps to perpetuate geographic maldistribution.[3-6] Since community characteristics and physician preferences will not change much in the near future, strategies to increase the supply of rural physicians should include methods for improving the conditions of rural practice.[3-8]

The large body of research on health manpower distribution suggests a number of factors that correlate with or influence physician location.[8-19] For people interested in attracting physicians into rural areas, these studies help to characterize probable candidates and appropriate inducements for rural practice. But why bother recruiting physicians for rural communities if they find rural practice conditions unacceptable and soon decide to leave? If the goal is to fill the basin, it seems only sensible to plug the drain first.

To understand why physicians quit rural practice, studies could ask physicians who

---

From the Robert Wood Johnson Clinical Scholars Program, University of North Carolina, Chapel Hill, North Carolina.

Address for reprints: Alan K. Kronhaus, M.D., Clinical Scholars Program, Box 5, Medical School Wing B 207H, University of North Carolina, Chapel Hill, NC 27514.

0025-7079/81/1200/1239/$00.90 © J. B. Lippincott Co.

once worked in that setting why they left. In the only published study of this sort, Parker and Sorenson mailed a closed-ended questionnaire to physicians who had once worked in rural areas of upper New York State. The 49 respondents mentioned inadequate time away from the practice as one of the most important reasons for leaving their rural communities.[3]

Surveys of physicians who switch from primary care to specialty practice can supplement the scanty data pertaining strictly to rural practice. The reasons physicians leave primary care bear on the issue of rural manpower because primary care physicians provide most of the services in rural areas.

From an open-ended questionnaire mailed to Virginia physicians who changed from primary care to specialty practice, Crawford and McCormack concluded that among their 73 respondents, "long hours" and the "strain of constantly feeling unable to get away" were the major reasons for dropping primary care practice.[4] Their conclusion was supported by the fact that 85 per cent of the physicians reported taking two weeks or less of vacation per year, which compares unfavorably with the national average of nearly five weeks of vacation per year for primary care physicians.[20]

In a series of personal interviews, Harvey asked 109 general practitioners in Canada why they thought physicians left general practice, and 74 physicians who had switched from general to specialty practice why they had changed. Both groups believed that "long hours" and "insufficient autonomy over one's time" were the most important factors influencing attrition in general practice.[5]

Attitudes and opinions do not establish cause and effect relationships, and all physicians do not react in the same way to the same conditions of practice. Nevertheless, three studies using very different methods strongly suggest that insufficient time off is a problem likely to evoke widespread dissatisfaction among physicians in rural practice and to aggravate attrition in rural areas.

Physicians could increase their free time in various ways. By giving patients adequate notice, they might simply take more time away from the practice. But many physicians hesitate to leave their practice without some type of coverage. Conversely, lack of suitable coverage might deter physicians from taking more time off. Group practice has been widely touted as a means to improve the conditions of rural practice, in part because a group can offer readily available coverage and more regular working hours to its members.[8,19]

The population of many rural communities, however, is not sufficient to support a group practice, or even another physician. Some physicians avoid making coverage arrangements with local colleagues because they fear losing patients, their standards of care do not coincide or they shun the added responsibility of providing coverage in return.

An organized service that provides locum tenens coverage in rural areas could benefit physicians and their patients in many ways. By helping physicians circumvent coverage problems, such an arrangement might increase their free time while improving continuity of care, maintaining practice productivity and assuring patient satisfaction during an absence. However, if physicians believe that locum tenens coverage would increase the expense of leaving their practice, economic concerns might overshadow other considerations. This study was undertaken to evaluate the economic productivity of locum tenens physicians in rural clinics and to compare the cost of leaving a practice with and without locum tenens coverage.

Setting

The data for this study were provided by Health Systems Research Institute (HSRI), a private corporation whose primary objective was to establish and operate medical clinics in manpower-shortage areas of the western United States. Medically underserved communities contracted with HSRI to recruit physicians who would make a long-term commitment to the community and its clinic. These physicians will be referred to as clinic physicians (CPs).

HSRI's strategy for recruiting and retaining clinic physicians was to provide them with a package of support services designed to improve the conditions of practice in their community. One element of this package was guaranteed time off with coverage. To provide the coverage, HSRI organized and deployed a group of locum tenens physicians (LTs). Some of these physicians entered into an exclusive agreement to do locum tenens work for HSRI and will be referred to as full-time locum tenens (F-LTs). The demand for coverage often exceeded the capacity of the F-LTs to provide the service. To obtain additional coverage, HSRI contracted with physicians who did locum tenens work only occasionally, perhaps during vacation from a training program or their own practice. These LTs will be referred to as part-time locum tenens (P-LTs).

Although some physicians in the study had specialty skills, they all functioned as generalists, and all did at least some obstetrics. Three CPs did a small amount of major surgery. Physician income was based on two different methods of compensation. CPs were guaranteed a minimum income, but could earn somewhat more by increasing their productivity. LTs were paid a fixed amount for each week of coverage, regardless of productivity.

The clinics were operated by HSRI and were required to maintain detailed work records for every physician. HSRI also provided coverage to clinics it did not operate, but these clinics were not included in the study because they did not send physician work records to HSRI.

The clinics did a minimum, if any, of their own laboratory or x-ray work; so the revenue generated represents billings for professional services. The clinics were located in communities that had populations of approximately 3,000–8,000 people and very small hospitals. Driving time to the next community with similar or more sophisticated services averaged 30–70 minutes under good conditions.

HSRI charged clinics a flat rate of $375/day for coverage. According to HSRI officials, the $375/day charge was enough to defray the direct (LT compensation and expenses) and the indirect (a realistic share of corporate overhead) costs of providing the service and to yield a reasonable profit for the company. The revenue billed by LT physicians was collected by the clinics to offset the amount charged for the service.

Methods

Productivity was expressed as the average daily revenue billed by LTs and CPs. The study compares LT productivity with CP productivity, and with the $375/day charged by HSRI for its LT service. It also compares the cost of taking time off with and without LT coverage.

Between January 1976 and June 1977, HSRI required its clinics to submit monthly records of the revenue billed and days worked by every physician. The records from the six clinics receiving coverage (one solo clinic, representing the work of two consecutive physicians due to turnover during the study period, and five partnerships) were available for examination at HSRI's central administrative office. A physician's average daily billing was calculated by dividing the revenue he or she billed by the number of days worked during each month of the study.

The 18-month study period encompassed the work of 12 CPs and 20 LTs. The locum tenens group was evenly divided between full-time and part-time physi-

TABLE 1. Comparison of Productivity: Locum Tenens with Clinic Physicians (N = 35 Episodes of Coverage)

	No. of Physicians	Average Daily Billings ($)	p Value (t-test)
CPs	12	$584	< 0.05
LTs	20	$439	

cians (F-LTs and P-LTs), who together provided 279 days of coverage, divided into 35 episodes varying in length from 3 days to 4 weeks. F-LTs accounted for 13 episodes, or 120 days of coverage, and P-LTs for 22 episodes, or 159 days of coverage. This represents the complete record of LT activities in all HSRI clinics receiving coverage during the study period. For each episode of coverage, LT productivity was compared with CP productivity during the same month.

Results

CPs billed an average of $584/day, compared with an average of $439/day for LTs. The difference is financially meaningful and statistically significant at less than the 1 per cent level (Table 1).

Figure 1 shows the relationship between locum tenens and clinic physician productivity. The ordinate indicates LT productivity during all 35 episodes of coverage, and the abscissa indicates CP productivity during the same month. There is no statistically significant correlation between the productivity of LTs and the productivity of CPs in the same practice, at the same time of year ($r = 0.23$, $p > 0.10$).

The distribution of LT billing in relation to the amount charged for their service helps to answer the question of how much it costs to take time off with LT coverage. HSRI charged the clinics an average of $375/day for LT coverage. Episodes of coverage during which LTs billed less than $375/day fall within the stippled area

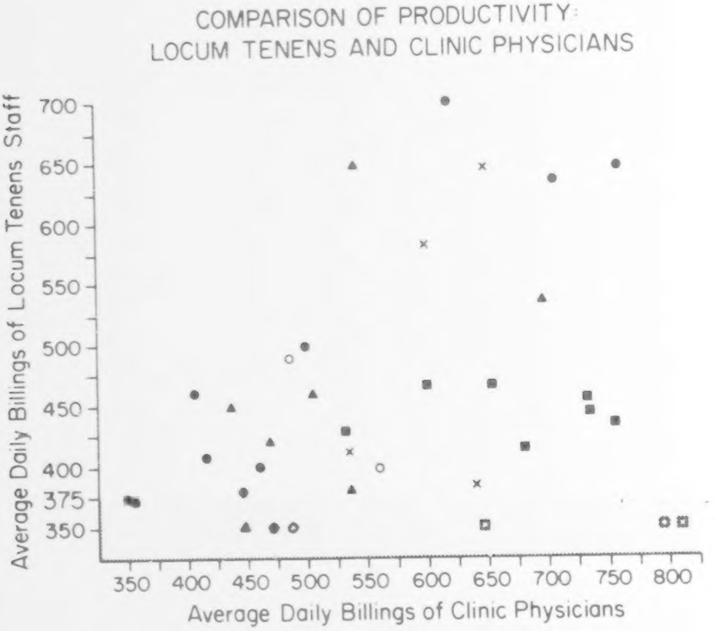


FIG. 1. Each point in the figure represents an episode of coverage. There are six different symbols, each representing one of the six clinics. The black circles represent the solo clinic. The stippled area designates episodes of coverage during which the locum tenens physician billed less than the amount charged for coverage.

TABLE 2. Comparison of Productivity Among Locum Tenens Physicians

	No. of Physicians	Episodes of Coverage	Average Daily Billings ($)	p Value (t-test)
Practice Experience				
> 10 years	7	11	$464	N.S.
< 10 years	13	24	$458	
Sex				
Male	18	30	$438	N.S.
Female	2	5	$468	
Length of Coverage				
> 2 wks.	10	11	$446	N.S.
< 2 wks.	15	24	$441	
Full-time vs. Part-time				
F-LT	10	21	$418	N.S.
P-LT	10	14	$479	

in Figure 1. LT billing during 28 of the 35 episodes (80 per cent) exceeded the amount charged for the service. From the perspective of individual physicians, as opposed to episodes of coverage, the average daily revenue billed by 17 of the 20 LTs (85 per cent) also exceeded $375/day.

Table 2 indicates the results of testing several independent variables for their effect on LT productivity. No statistically significant relationship was observed for practice experience, sex, duration of coverage or the LTs' status as full-time or part-time members of the group.

Another way to gauge the productivity of LTs is by their ability to maintain a CP's customary level of billing. When LT billing is expressed as a percentage of CP billing during the same month in each clinic, LT productivity was found to be significantly higher in solo than in partnership practices (97 per cent vs. 73 per cent, $p < 0.01$; see Table 3).

In order to compare the cost of leaving a practice with and without LT coverage, the assumption was made that practice overhead remains constant when the CP takes time off. If overhead stays the same, the cost of leaving the practice without coverage is approximately equal to the revenue forgone during the physician's absence. For CPs in this study, the average cost of leaving the practice without coverage would have been $584/day. By paying HSRI $375/day for LTs who billed $439/day, the clinics could realize a profit of $64/day from their investment in the service—if they collected the entire amount. On the average, LTs reduced the cost of an absence by 11 per cent, down to $520/day (see Table 4).

Discussion

An organized locum tenens service offers important advantages both to patients and to physicians unwilling or unable to arrange coverage locally. During an absence, the service could improve con-

Table 3. Comparison of Productivity: Locum Tenens in Solo and Partnership Practices

	Episodes of Coverage (N = 35)	% of CP Billing Achieved by LT	P Value (rank sum test of 35 episodes)
Solo	11	97	< 0.01
2-MD	24	73	

TABLE 4. Cost of Leaving with and without Coverage, in Dollars per Day

	Without Coverage	With Coverage
Billing during absence ($)	None	439
Revenue forgone during absence ($)	584	584 − 439 = 145
Direct cost of LT service ($)	None	375
Net cost of absence ($)	584	145 + 375 = 520

tinuity of care, maintain practice productivity, prevent practice dissipation and assure patient satisfaction. If the service could increase free time for rural physicians and thereby improve the conditions of rural practice, it might eventually help to stabilize or increase the supply of rural manpower by making physician recruitment easier and retention more certain.

But will physicians be receptive to this locum tenens arrangement? The answer depends on both the quality of the service and the cost. This study demonstrates that it will most probably cost a physician less to leave his or her practice with LT coverage than without coverage. So from a strictly economic point of view, the LT arrangement should be well received.

The conclusion that the cost of leaving a practice with LT coverage is less than the cost of leaving without coverage is based on two assumptions: revenue forgone during absence approximates the revenue physicians would have produced had they been working during that period, and overhead costs do not change when physicians leave. If it were feasible and desirable, physicians could reduce the cost of absence in two ways. They could reduce revenue losses during absence by increasing patient care activity before or after they leave, or they could cut overhead costs by decreasing staff.

There are other factors that would weigh in a more thorough accounting of the financial impact of an LT service on a physician's practice. Preservation of a CP's hospital census, protection from patient attrition through competition during an absence and the fact that the LT service is tax deductible as a business expense are factors that would shift the accounting in favor of the LT arrangement. The revenue generated by laboratory and x-ray facilities, which were virtually absent from the study clinics, should increase the probability of an LT billing more than the amount charged for his or her service. On the other hand, a poor collection rate would hamper a physician's ability to recover the cost of the service.

In some respects, collected revenue might have been a better basis for determining how much the service cost a CP. There are problems with using that variable, however, especially in the comparisons of physician productivity. The nature of clinic accounting systems makes it virtually impossible to attribute collected revenues to a CP or LT, even in the solo practices, because patient accounts often add up over time, people make partial payments against their account balance, and so forth. Moreover, the collection rate was largely a function of HSRI's skills and aggressiveness in pursuing accounts, so that "collections" reflect activities other than physician productivity.

Half of HSRI's LT group was somewhat unique because they did full-time locum tenens work for HSRI. Their productivity, however, did not differ from the productivity of the part-time LTs who may be more representative of physicians likely to be engaged in locum tenens work. It would be interesting to compare the productivity of the LTs associated with HSRI with the productivity of physicians who do locum tenens on their own, or of physicians who do locum tenens for the American Medical Association's Project U.S.A. That program relies on physicians, often semiretired, who occasionally provide locum tenens coverage to government-supported clinics.

It is no surprise that LTs billed less than CPs. Unless there were excess demand for physician services in the area, some patients would surely prefer to postpone a visit until their regular physician returned. This rescheduling would decrease LT utilization and productivity, with a reciprocal increase in CP productivity. The finding that LT productivity more closely matched CP productivity in solo than in partnership practices may also be due to in part to patient preference. In the partnerships, one CP had to remain in the practice when the partner left, so patients of the absent physician had a convenient and more familiar person they could choose in preference to the LT.

The difference between CP and LT productivity might also be explained in part by differences in the way they were compensated. LTs were paid a fixed amount for each week of coverage, regardless of productivity; CPs were guaranteed a salary, but could earn somewhat more by increasing their productivity. This difference, plus the fact that LTs worked for HSRI and not directly for the clinics, may have blunted their motivation to maintain the same productivity standards as the clinic physicians.

It is possible that LT productivity might have been higher in HSRI clinics than it would be in other rural practices. Patients might have visited the LTs more readily because HSRI communities were quite distant from the next available source of care. Clinic personnel can promote or discourage utilization of an LT. HSRI encouraged its employees to maintain normal scheduling during periods of coverage, a practice that would contribute to LT productivity.

Another factor that could have influenced LT productivity in HSRI communities is their long history of physician turnover. People inured to seeing different physicians might accept an LT more readily then people accustomed to seeing the same physician.

Many small, isolated communities have experienced longstanding difficulties in securing a physician. An organized LT service might offer these communities more stable health care by providing continuous physician coverage on a rotating basis. LT physicians could also provide seasonal coverage in national parks, resorts or other tourist areas.

This study suggests that rural physicians can benefit financially from locum tenens coverage. Future research could broaden the scope of this study to include specialties other than family practice and other important criteria for evaluating a locum tenens service. These criteria might include its effect on patient as well as physician satisfaction, the quality and continuity of care, physician attrition in rural areas and the cost of care to patients and society.

Acknowledgments

The author is deeply indebted to Robert Fletcher, M.D., and to the Robert Wood Johnson Clinical Scholars at the University of North Carolina, Chapel Hill, for their generous support and guidance; and to Edward Wagner, M.D., Jo Anne Earp, Sc.D., Gordon DeFreise, Ph.D., Dennis Gillings, Ph.D., Deborah Freund, Ph.D., and Brett Williams, M.D., for their constructive assistance during the preparation of this study. Therus Kolff, M.D., of Comprehensive Health Systems (formerly Health Systems Research Institute), was kind enough to provide access to the data used in this study.

References

1. Ginzberg E, Ostow M. Men, money, and medicine. New York: Columbia University Press, 1969:113.

2. Fuchs VR. Who shall live. New York: Basic Books, Inc., 1974:69.

3. Parker RC, Sorenson AA. The tides of rural physicians: the ebb and flow, or why physicians move out of and into small communities. Med Care 1978;16:152.

4. Crawford RL, McCormack RC. Reasons physicians leave primary practice. J Med Educ 1971;46:263.

5. Harvey EB. The vanishing practitioner. J Med Educ 1973;48:718.

6. Rakel RE. A better break for the rural physician. Prism 1973;1:56.

7. Kronhaus A. Rural practice reconsidered—a method of organizing health resources. Clin Med 1976;83:19.

8. Madison DL. Recruiting physicians for rural practice. Health Services Rep 1973;8:758.

9. Parker RC, Tuxill TG. The attitude of physicians toward small community practice. J Med Educ 1967;42:327.

10. Parker RC, Rix RA, Tuxill TG. Social, economic, and demographic factors affecting physician populations in upstate New York. NY State J Med 1969;69:706.

11. Taylor M, Dickman W, Kane R. Medical students' attitudes towards rural practice. J Med Educ 1973;48:885.

12. Martin ED, Moffat RE, Falter RT, Walker JD. Where graduates go: The University of Kansas School of Medicine—a study of the profile of 959 graduates and the factors which influenced their geographic distribution. J Kans Med Soc 1968;69:84.

13. Bible BL. Physicians' views of medical practice in nonmetropolitan communities. Public Health Rep 1970;85:11.

14. Cooper JK, Heald K, Samuels M, Coleman S. Rural or urban practice? (an analysis of factors influencing the primary care physician's practice location). Inquiry 1975;12:18.

15. Hassinger EW. Background and community orientation of rural physicians compared with metropolitan physicians in Missouri. Research Bulletin. University of Missouri, College of Agriculture, 1963:822.

16. Peterson GR. A comparison of selected professional and social characteristics of urban and rural physicians in Iowa. Health Care Research Series 8. University of Iowa, 1968.

17. Champion DJ, Olsen DB. Physician behavior in southern Appalachia: some recruitment factors. J Health Soc Behav 1971;12:245.

18. Eisenberg BS, Cantwell JR. Policies to influence the spacial distribution of physicians. Med Care 1976;14:455.

19. Evashwick CJ. The role of group practice in the distribution of physicians in nonmetropolitan areas. Med Care 1976;14:808.

20. The American Medical Association. Profiles of medical practice, 1978:192.

-1 c 5170-1

SUPPLEMENT TO

Medical Care

December 1981 · Vol. 19 No. 12
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Access to Medical Care: Progress, Problems and Prospects

JAMES LOGERFO, M.D., M.P.H., SUPPLEMENT EDITOR

PUBLISHED BY
J. B. Lippincott Company · Philadelphia · Toronto

CONTENTS

Volume 19 December, 1981 Number 12, Supplement

Supplement Editor's Note ii

Acknowledgment iii

Introduction. *Susan Morgan* 1

Equity of Access to Medical Care: A Conceptual and Empirical Overview. *Lu Ann Aday and Ronald M. Andersen* 4

Equity in the Distribution of Quality of Care. *Leon Wyszewianski and Avedis Donabedian* 28

Health Status as a Measure of Need for Medical Care: A Critique. *John Yergan, James LoGerfo, Stephen Shortell, Marilyn Bergner, Paula Diehr and William Richardson* 57

Equity, Access, and the Costs of Health Services. *Bruce C. Vladeck* 69

MEDICAL CARE
December 1981, Vol. XIX, No. 12, Supplement

Equity of Access to Medical Care: A Conceptual and Empirical Overview

LU ANN ADAY, PH.D., AND RONALD M. ANDERSEN, PH.D.

This paper summarizes the major ethical considerations in the conceptualization and measurement of the equity of access concept, presents national and community data on the most current profile of access in the United States, and discusses the implications of these conceptual and empirical issues. Changes in health care policy and the future of the "equity of access" objective in the U.S. are also discussed.

THE ETHICAL considerations in the conceptualization and measurement of equity of access to medical care, the current empirical profile of equity in the United States, and the implications of these conceptual and empirical issues for future health policy dealing with the "equity of access" objective are the focus points of this report.

Implicit in this objective of equity are three assumptions that serve as a starting point for examining the ethics of equity: 1) Health care is a right; 2) The resources for allocating health care are finite; 3) Health policy should be concerned with the design of "just" mechanisms for allocating scarce health care resources. These assumptions form the basis for much of the conceptual and methodological interest in equity. Selected literature dealing with these assumptions is summarized to provide an overview of the normative context from which fundamental questions regarding the "ethics" of the equity concept have emerged in recent years.

Right to Health Care

In the medical ethics literature, a great deal of attention has been devoted to examining the concept of the "right to health." Blackstone,[1] for example, explores the constitutional and moral bases for this right in terms of equal protection, minimum welfare, and other "justice" criteria. Sidel[2] and White,[3] however, point out the importance of distinguishing the "right to *health*" from the "right to *health care*." They argue that a variety of factors may contribute to health that are beyond the power of the state to ensure it as a basic "right," e.g., personal behavior that leads to ill health. The "right to health care," however, implies *access* to medical care services, which may or may not lead to "health" per se. Carlson[4] and Illich[5] argue, for example, that the modern medical care system may in fact be ineffec-

---

Center for Health Administration Studies, Graduate School of Business, The University of Chicago, Chicago, Illinois.

The Community Hospital Program (CHP) access data are subject to change.

This paper was prepared for the President's Commission for the Study of Ethical Problems in Medicine and Biomedical Behavioral Research. The original research on which some of the data reported here were based was supported by grants from The Robert Wood Johnson Foundation (Princeton, New Jersey) and Contract #HRA 230-76-0096 from the National Center for Health Services Research (Hyattsville, Maryland).

Address for reprints: Lu Ann Aday, Ph.D., Center for Health Administration Studies, The University of Chicago, 5720 S. Woodlawn Ave., Chicago, IL 60637.

tual or even deleterious in effecting "health" in patients who use it.

Blumstein and Zubkoff,[6] Mechanic,[7] and Sparer,[8] however, point out that in general "the right to health care" is concerned with ensuring that everyone has a right of access to services and that this right is an important normative goal of the health care system. McGarvey[9] further distinguishes the "right *to* health care" from the "right *in* health care" in that where the former is concerned with "access," the latter deals with the character of the individual interaction between the patient and the physician, once entry (access) is gained. Both are or should be proper concerns of medical ethics, but it is the right *to* medical care that is one of the most fundamental *normative* starting points for the consideration of access per se.

Finitude of Health Care Resources

Gustafson[10] argues that when the finitude of resources to satisfy human needs is realized, one's own self-interest is secured or greater attention is paid to applying principles of distributive justice for the common good. The choice between these alternatives is, in a sense, a moral one. The shortage of a desired good relative to the demand for it is expressed in the economic marketplace through high prices for that scarce resource. The finitude of resources for securing the right to health care is correspondingly visible in the rising cost of medical care, borne by individuals and society, as the demand for and price of high-technology care continue to increase.[11–13]

The recognition of the finitude of health care resources and the means for allocating them serves to establish "real world" parameters on how or whether the "moral" right or goal of access to medical care can actually be realized. Further, to the extent resources are devoted to the realization of the access goal, other health care system goals such as cost containment and quality may well be compromised. These tradeoffs are also of essential concern in understanding the ethics of health policy-making overall.

"Just" Allocation of Finite Health Care Resources

Much of the debate concerning means for assuring the "right to health care" concerns whether health care differs from other commodities purchased and consumed in the marketplace, the applicability of the "market" model to distributing health care, and the "fairness" of the various mechanisms for allocating finite health care resources.[7,12,14–16]

At the heart of considering the concept of equity of access is the question: What criteria define a "just" system of allocating this scarce resource? The relevance of variant criteria to current conceptual and methodological approaches to evaluating equity is the focus of the following section.

Conceptual and Methodological Issues in Equity of Access

A variety of approaches to defining and measuring access to medical care have appeared in the literature in recent years. Some researchers have emphasized the overall availability (supply) of services; others have focused on the characteristics of the people who may potentially avail themselves of these services, considering factors such as their income levels or insurance coverage. More recently the focus of access studies has shifted to intermediate outcome measures, such as the rates at which services are actually used or how satisfactory consumers perceive their care to be.

The authors have developed a framework for studying access to medical care that attempts to integrate these diverse approaches.[17,18] In this framework, access is defined as "those dimensions which describe the potential and actual entry of a given population group to the health care delivery system."[19] Figure 1 shows an outline of the elements of this framework.

Further, in this model *equity* of access to care is said to exist "when services are distributed on the basis of people's need for them . . . Inequity is suggested, however, if services are distributed on the basis of demographic variables, such as race, family income, or place of residence, rather than need."[19] In terms of Figure 1, variation in the use of services as a function of the need component or associated age and sex correlates within the predisposing variables represent equity. Variation that is a function of the availability of services or how they are organized or other predisposing or enabling characteristics of the individuals themselves means that services are not equitably distributed.

In the discussion that follows, alternative approaches to measuring equity are introduced with particular reference to how they might be put into operation within the access framework outlined in

FIGURE 1. Framework for the Study of Access

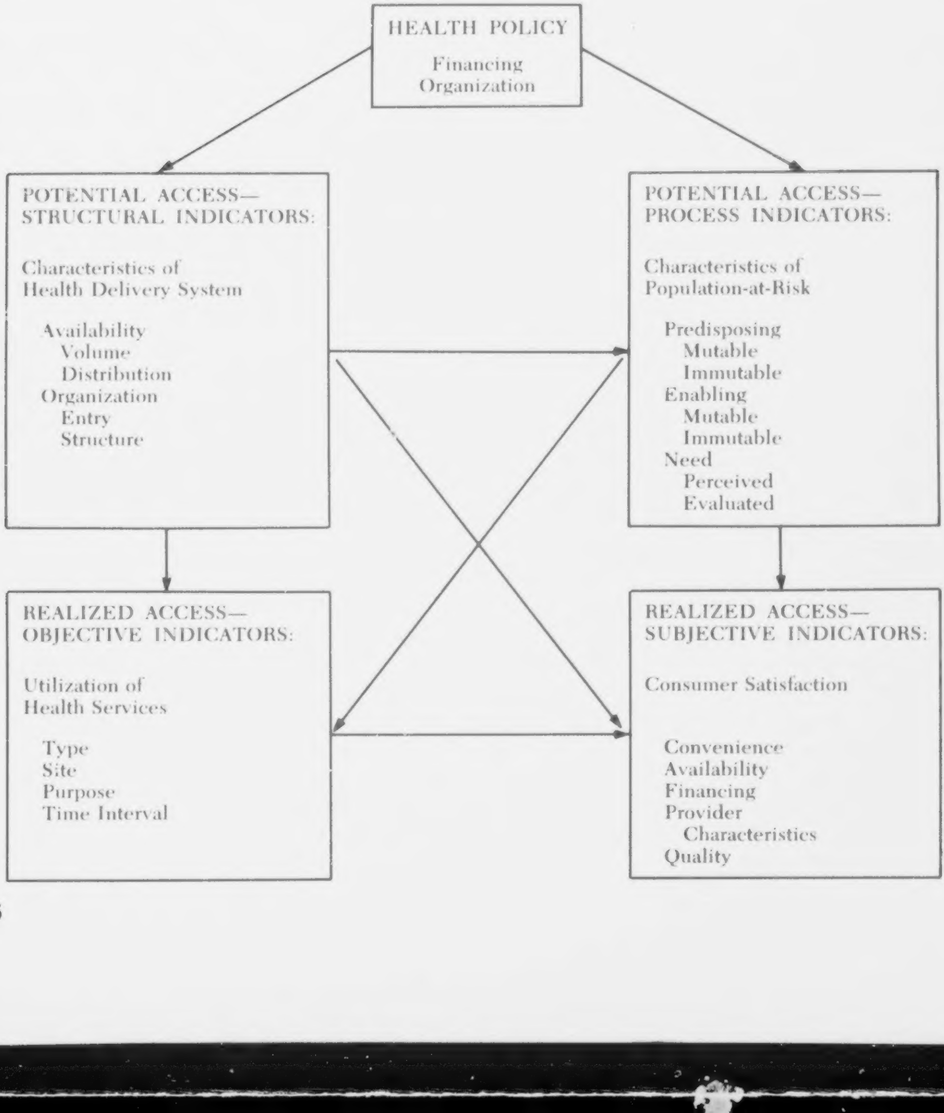

HEALTH POLICY
Financing
Organization
POTENTIAL ACCESS—STRUCTURAL INDICATORS:
Characteristics of Health Delivery System
Availability
Volume
Distribution
Organization
Entry
Structure
POTENTIAL ACCESS—PROCESS INDICATORS:
Characteristics of Population-at-Risk
Predisposing
Mutable
Immutable
Enabling
Mutable
Immutable
Need
Perceived
Evaluated
REALIZED ACCESS—OBJECTIVE INDICATORS:
Utilization of Health Services
Type
Site
Purpose
Time Interval
REALIZED ACCESS—SUBJECTIVE INDICATORS:
Consumer Satisfaction
Convenience
Availability
Financing
Provider Characteristics
Quality

Figure 1 and the applicability of these respective criteria for effecting a "just" allocation of health care resources.

Concepts of Equity

Outka[20] explores the relevance of a number of criteria of social justice to the "equal access" objective. They are: 1) merit or desert, 2) societal contributions, 3) contribution to satisfying whatever is freely desired by others in the open marketplace, 4) needs, and 5) similar treatment for similar cases. These concepts serve as the starting point when considering variant criteria for measuring equity and the ways in which other conceptions or interpretations of justice modify or extend these.

Merit. Outka argues that merit, or who is *deserving* of care, is not an appropriate criterion to apply when considering "just" means for allocating health care resources because 1) the idea of justice is not fully defined by the merit criterion alone, and 2) health crises generally cannot be based on merit because they occur beyond an individual's power to control or predict them.

Outka does describe examples in which it can be argued that individuals are to some degree responsible for their medical problems, e.g., an overweight person who had a heart attack, a man with lung cancer who smoked cigarettes for over 40 years, or a football player who suffered a concussion.

A meaningful application of this merit criterion in terms of the overall access objective, however, may be to examine those predisposing factors, such as attitudes or health behaviors, that place individuals at greater risk of needing health care. Examples of these are people with the smoking or exercise habits in the previous example, as well as persons in age and sex categories for which the biologic givens of aging or anatomy predispose them to "undeserved" health risks. These criteria are, in fact, applied in health education, screening, risk evaluation programs, and in insurance policies with reduced premiums for people with good health habits (those who do not smoke or drink, for example).

Societal Contribution. The societal contribution criterion is based on a utilitarian concept of justice that considers the greatest good for the greatest number. Outka points out that those who contribute least to the general welfare (such as the elderly or disabled) may be ones for whom access to health care support services are the most critical.

In our society, occupational and educational statuses are the criteria frequently used (either implicitly or explicitly) to assess whether an individual's contribution to the common good is worthwhile. The differential distribution of employment-related health care benefits, for example, may be viewed as a manifestation of the resources that society is willing to devote to sustaining the health of differentially valued societal "contributors." Professional and occupational statuses cannot entirely determine whose "right to health" is most fully guaranteed, however. Medicaid and Medicare, for example, were enacted to provide coverage for individuals whose contributions to society may well be the least—the poor and the elderly.

Supply and Demand. The supply and demand principle of resource allocation emphasizes the consumer's free choice of preference among alternatives and the distribution of services on the basis of the unrestricted (laissez faire) operation of these forces. Outka and others, however, argue that the application of these principles to the distribution of health care resources is inappropriate primarily because the consumer has much less choice when deciding whether to purchase this type of service. Outka, for example, points out that "when lumps appear on someone's neck, it usually makes little sense to talk of choosing whether to buy a doctor's service rather than a color television set."[20] Mechanic[7,12] also argues that physicians, rather than consumers, make the major decisions about which services should be purchased.

The consumer has little information or knowledge of the appropriate "commodity" to buy when he or she is seriously ill. Navarro,[15,21,22] in fact, has made the case that the health insurance industry and other powerful corporate interests, rather than either physicians or consumers, make the major decisions about the kinds of health care to be delivered and consumed in the health care marketplace.

Economists are most apt to be proponents of the unrestricted operation of the free market forces of supply and demand for the allocation of health care resources.[23,24] Programs should be designed, they argue, to encourage rather than restrict the consumer and the physician when deciding on the amount and price of health care.[14,25] This policy perspective tends to encourage the development of competitive models of health care delivery, such as HMOs, or the provision of income maintenance rather than health-care specific subsidies. The laissez faire principles in this approach conform closely to Nozick's entitlement concept of justice, in which free choice among alternatives and entitlement based upon those historic choices are the main criteria on which a just allocation of resources is based.[26]

The access framework outlined earlier summarizes the supply (delivery system characteristics) and demand (population characteristics) factors that determine the price and quantity of services consumed in the health-care marketplace. A key health policy concern in examining these forces or influences is the extent to which there should be intervention to directly regulate the behavior or operation of these determinants of who obtains how much care. Most contemporary medical economists argue that some "basic minimum" care should be assured by society; however, market forces should be freer to operate in allocating services beyond this minimum.[6,25] Fein[26] argues for introducing new incentives into the existing market (e.g., centralized ceiling on health expenditures, prepaid forms of organizing care) and allowing the system to respond to these incentives.

Need. Outka[20] contends that "essential needs," in fact, constitute a much more appropriate basis for the allocation of health care resources than do merit, societal contribution, or supply and demand criteria alone. The previous discussion has suggested that because of the ethical uniqueness of allocation decisions concerning medical care, these latter criteria have, in fact, either been tempered, transcended, or ignored in many major health policymaking efforts.

As mentioned earlier, this criterion based on need has been the primary one used by the authors to describe an equitable system of health care resource distribution . Rawls' concept of social justice, based upon the allocation of resources to those individuals in the society who have the greatest need, lends support to this approach, as do those who contend that medical care is a unique commodity for which the typical market force of merit-based allocative criteria may not be appropriate.[27,28] For example, Fein has argued that we should "reject market mechanisms that might allocate resources to and within the health sector, in part, because market results are at variance with our values."[26]

The need criterion may be made operational in terms of a minimum service (e.g., annual Pap smears for women in childbearing years), the identification of pockets of observed morbidity (e.g., lead paint poisoning), or measures of the fit between observed need and obtained services (e.g., use of services relative to experienced symptoms or disability). It is this latter utilization criterion based on need that has received the greatest attention in our own approach to measuring equity.

Implicit in the development of these need criteria are some assumptions regarding the appropriate level of care relative to experienced need. The development of these criteria often requires consumers or providers of care to make value judgments on how much of what service is required.[6] In addition, methodological criticisms have been made of the techniques used to put this approach into operation.[19] Nonetheless, it is a perspective that has received considerable attention as a basis for allocating health care resources and will be of particular interest in some of the empirical analyses of access that follow.

Similar Treatment for Similar Cases. This criterion can be explored in the context of the access framework described earlier when differences in the rates at which people report having a regular source of medical care, the convenience of this source, or the satisfaction with this care are examined. Income, race, or whether a person is covered by Medicaid or private insurance should not be used to determine the medical care received by people who must enter different doors or be "treated" differently by the health care system. This criterion, in particular, focuses on sociodemographic subgroup comparisons. Fein refers to distributional equity on this basis as "vertical equity."[26]

Measurement of Equity

The preceding discussion has described the primary ways in which the equity of access concept has been defined when formulating health policy decisions. Table 1 summarizes how each of these approaches can be put into operation, with particular reference to the access framework outlined earlier.

In our application of the access framework to measure the equity concept, we have tended to focus on the last two concepts of equity that are based on need and similar treatment for similar cases. In particular, we have emphasized the development of need-based measures of utilization and comparisons of race, income, residence, and other groups for whom similar treatment for comparable levels of need are expected.[19]

As Table 1 suggests, however, the access framework can actually be applied when putting into operation each of the approaches used to measure equity just described. Personnel and facility-to-population ratios, for example, are useful in exploring the operation of supply and demand factors. The convenience and characteristics of places people go for medical care provide data on whether there is differential "treatment" of individuals, depending upon where they happen to go for services.

Age is a correlate of need as well as a basis for defining actual or potential societal contributions; e.g., children, working-age adults, and the elderly. Sex, similarly, is often considered, rather than need; race serves to identify groups for whom access differences may well indicate differential treatment. As mentioned earlier, education and occupation are used as proxy indicators of the contributions individuals make to society. Health-care attitudes and behaviors reflect predispositions toward health and health care that may serve as a basis for deciding whether certain individuals merit the allocation of scarce health care resources. Health habits, such as smoking, diet, and exercise, did not appear in our initial access framework. They are, however, ultimately important determinants of the need for and utilization of health-care resources and may be appropriately considered among the factors predisposing individuals to eventually seek services.

The rural-urban distribution of services is important from the point of view of both the supply and demand and similar treatment concepts of equity. According to the similar treatment norm, for example, where a person lives should not determine the adequacy of medical services available. Family income serves as an indicator of the "value" individuals have in society, as well as a basis for determining the operation of supply and demand factors. Differential access by family income, for example, may

TABLE 1. Application of Access Framework to the Measurement of Equity Concepts

		Concepts of Equity				
Dimensions of Access	Empirical Indicators of Access	Merit	Societal Contribution	Supply and Demand	Need	Similar Treatment for Similar Cases
Characteristics of health delivery system						
Availability	Personnel/facility-to-population ratios			X		
Organization	Structure and convenience of medical care facilities					X
Characteristics of population-at-risk						
Predisposing	Age		X		X	
	Sex				X	
	Race					X
	Education		X			
	Occupation		X			
	Health care attitudes	X				
	Health care behaviors*	X				
Enabling	Rural-urban place of residence			X		X
	Family income		X	X		X
	Structure and convenience of regular source of care					X
	Type and extent of insurance coverage		X	X		
Need	Disability days	X			X	X
	Symptoms	X			X	X
Utilization of health services						
Physician				X	X	X
Hospital				X	X	X
Dentist				X	X	X
Consumer satisfaction						
General satisfaction						X
Visit-specific satisfaction						X

* This variable did not appear in the original access framework.

suggest that differential purchasing power rather than need per se determines who receives care.

Insurance coverage is a mechanism for effecting the operation of supply and demand factors and ultimately the price of services to those whose societal contribution is great (e.g., employees) or whose previous (elderly) or potential (children) societal contributions merit attention. The characteristics and convenience of places people usually go for medical care indicate whether individuals seem to have different experiences with different types of providers (e.g., longer waiting times in hospital emergency rooms, shorter visits with general practitioners compared with specialists) and what these differences might mean in terms of the "equitable" allocation of health care resources.

Disability days and symptoms are empirical indicators of underlying medical morbidity. They may be used as statistical controls to determine whether access differentials persist. Morbidity is also important in determining whether certain individuals merit the allocation of health care resources (e.g., life-threatening or potentially disabling illnesses may signal those conditions most deserving of attention).

Utilization measures are essential in supply and demand formulations of equity. Utilization indexes based on need are also important when need is used as the basis for resource allocation. That is, we expect similar utilization rates for individuals whose needs are similar.

Satisfaction levels indicate whether people have experienced differential treatment. In general, the more inconvenient or costly people's experiences with the health care system, the less satisfied they are.

The operational indicators just reviewed are the ones that are oftentimes available from empirical studies of access.[29] They are more or less relevant, as indicated by the previous discussion, to measuring a particular concept of equity. Further, one indicator may be applied in more than one approach to the equity concept.

Methodological Problems in Measuring Equity

A number of methodologic problems that plague any effort to empirically measure equity should be noted, however. They are, in fact, limitations of health services research methodology in general. Their particular relevance to the measurement of the equity of access objective is summarized here.

Empirical research has failed, for example, to demonstrate a necessary relationship between gaining access to medical care and health. Many factors (e.g., genetics, environmental pollution, health habits, occupational stresses) other than health care per se can affect one's overall health status. There is some evidence, as cited earlier, that health "care" may, in fact, lead to poorer rather than better health. The equity of access objective, however, is concerned with individuals gaining the right to "take the chance" that obtaining such care will benefit rather than harm them.

The "science" of medicine has not been able to establish universally accepted criteria for what is the most appropriate care for any particular illness. Standards of the appropriateness of access for a given level of need are still in large measure judgmental. Some critics have argued, in fact, that consumers may be as equally qualified as medical professionals to judge whether it is appropriate for them to seek scientific medical care for particular illnesses. Consumer satisfaction levels and attitudes are equally relevant criteria for evaluating the "success" of the medical profession in achieving their curing and caring objectives.[30]

As Mechanic[31] and others have pointed out, much of the variance in understanding the utilization of or access to medical care services remains unexplained by large-scale empirical studies in this area. One may argue that empirical work to date has done a poor job of testing and measuring the equity of access concepts or that current conceptions do not adequately capture the full range of factors that may influence who gets care and why. The extent to which health services research can help determine health policy and the ethical choices associated with it is bounded by the success with which it can empirically model these choices and their implications for the individuals and the health-care system.[32] To the extent a great deal of the variance in access remains unexplained, the precise profile of equity may correspondingly be inadequately described.

Other methodologic issues used to empirically measure equity concern the direction and magnitude of observed relationships. For example, does not having a regular source of medical care mean people seek services less often, or do people who are healthier have less need to establish a relationship with a regular provider? Or how much difference by income or race is "enough" to establish that there is an access "problem?" Also, might the ethical importance of certain determinants vary depending upon the particular type of service being considered, e.g., emergency medical care versus cosmetic surgery. Further, Sloan[33] points out that different generalizations about equity may be the result of the particular concept or empirical definition of equity that is used.

Methodologic problems, therefore, do limit the generalizations that can be made about the profile of equity that exists at any point in time. In the section that follows the most current empirical information on the status of the equity goal is summarized, taking into account some of the limitations just described.

Trends in the Profile of Equity

Since the late 1950s, national data on the profile of access to medical care have shown increasing improvement for traditionally disadvantaged groups.[19,34,35] The advent of large-scale public financing programs (Medicare and Medicaid) and programs targeted to low-income populations (community health centers, children and youth programs, maternal and infant care) have been credited with these improvements.

Since the mid-1970s, however, there have been substantial changes in the overall economic and political climate of the country. Accompanied by programmatic changes, they may well temper the progress toward equity observed in previous decades.

We have witnessed in recent years, however, an era of double-digit inflation in tandem with high rates of unemployment. Tax payers have begun to more vigorously protest the amounts they must pay into the public coffers. Local, state, and federal lawmakers see cost containment and balanced budgets as increasingly important, but difficult, goals to achieve.

There is an increasing skepticism about what "big government" can accomplish and more blatant criticism of the perceived self-interest and inefficacy or even incompetence of the medical as well as other (e.g., legal) professions.

The poor economic health of the nation and general conservative attitude toward the efficacy and efficiency of existing programs have certainly had an impact on the health sector, as manifested by cutbacks in the Medicaid programs in a number of states, the closings of inner-city public hospitals in many large cities, the inability of Congress to reach agreement on a national health insurance proposal, and the increased burden on individual families who lose health insurance benefits during periods of unemployment.

Under these conditions, optimization of the goals of equity of access, cost containment, and quality is difficult to attain. There is some evidence that the public financing programs credited with substantial improvements in access for low income persons and nonwhites may have, in fact, significantly contributed to the escalation in the cost of medical care in recent years. Further, though utilization rates may be higher now for these groups than they ever were in the past, where they go for care (e.g., hospital emergency rooms, Medicaid mills, physicians who are not specialty-board certified) suggests that the quality of the care available to them may be less than optimum or, in fact, poor in many cases.

In the following section the most recent national data on the equity of access profile are presented in an effort to determine where we now stand. In particular, we examine whether the generally positive trends of the previous decades are continuing, whether they are apt to be maintained, and in which areas existent gains are most apt to be compromised as a result of the changes in the overall economic, political, and programmatic climate of the nation just described.

Current Survey Data on Equity

The primary sources of the data presented here are the national surveys of medical care utilization conducted by the Center for Health Administration Studies and the National Opinion Research Center (CHAS-NORC) at The University of Chicago,[19,34] data from the National Center for Health Statistics Health Interview Surveys (NCHS-HIS), and preliminary data from community survey evaluations of The Robert Wood Johnson Foundation Community Hospital Program (RWJF-CHP) carried out by the Center for Health Administration Studies in 1978 and 1979.[36,37]

The concepts of equity that were discussed earlier and are particularly emphasized in the data presented here are the need and similar-treatment-for-similar-case criteria. Poor–nonpoor and age-group comparisons are especially emphasized, as are adjustments for variants in need for services. Data are presented on both indicators of potential (e.g., insurance coverage) and realized (e.g., utilization) access to services.

Table 2 summarizes trends for five access indicators: per cent having third party (public or private) coverage, per cent of women with live births who saw a doctor during the first trimester of pregnancy, hospital admissions per 100 persons per year, doctor visits per person per year, and the per cent seeing a dentist in the year. Data are presented for five different points in time over an approximately 25-year period—1953, 1958, 1963, 1970, and 1976—covered by the CHAS-NORC national surveys.

The table shows the ratio of each indicator value for the poor compared with the nonpoor. A score greater than one means that the poor are in a better position (have higher scores) than the nonpoor. A value of less than one means they are in a less favorable position. A value of 1.00 means 100 per cent parity (or agreement on the actual scores) between the two groups.

Most broadly conceived, a value of one on the scale score axis indicates a situation of equity by the similar-treatment criterion. Some variance from this norm may well be due to real differences in need, which is a justifiable basis (according to the need criterion) for discrepancies among groups for certain of the indicators examined here (e.g., hospital admissions, physician visits).

According to Table 2 the poor continue to be at a relative disadvantage compared with the nonpoor with respect to the proportion having some form of third-party insurance coverage. Though with the advent of Medicare and Medicaid their situation has improved considerably since the mid-1960s, the proportion of the poor having any form of coverage (public or private) is still less than the nonpoor. Further, in recent years the poor do not appear to be gaining on the rest of the population.

The proportion of low-income women seeing a doctor during the first trimester of pregnancy continues to be lower than that of higher income mothers. There is evidence of substantial gains between 1953 and 1958, which have not been fully maintained in the past two decades. There was little change between 1970 and 1976 in the relative position of poor mothers. They continued at 73–74 per cent of parity compared with high income mothers.

The poor have generally had higher rates of hospital admissions than the nonpoor over the past 25 years. One reason is that

TABLE 2. Selected Indicators of Access to Medical Care of the Poor* Compared With the Nonpoor: From the CHAS National Surveys† (1953, 1958, 1963, 1970, 1976)

Access Indicators	Ratio Score: Poor/Nonpoor‡				
	1953	1958	1963	1970	1976
Per cent with third-party coverage of medical expenses	0.70	0.65	0.73	0.87	0.84
Per cent of women with live births who see physician in the first trimester of pregnancy	0.56	0.83	0.67	0.74	0.73
Hospital admissions per 100 persons per year	1.02	1.08	1.05	1.36	1.29
Physician visits per person per year (excludes phone calls)	0.70	—	0.98	1.32	1.12
Per cent seeing dentist within a year	0.37	—	0.48	0.57	0.60

* Source: References 19, 34, 38, 39, 40, 41.

† Definitions of poor based on total family income of under: (1953) $3,000; (1958) $3,500; (1963) $4,000; (1970) $6,000; (1976) $8,000.

‡ Score less than 1.00 indicates poor have less access; score greater than 1.00 indicates poor have greater access.

the poor tend to have much more serious, especially chronic, illnesses compared with the nonpoor. A major increase in the relative number of admissions of the poor compared with the nonpoor occurred in the post-Medicare-Medicaid period (from 5 per cent more in 1963 to 36 per cent more in 1970). This relative increase leveled off over the next 5–6 years, however.

Especially dramatic gains of the poor relative to the nonpoor may be observed in terms of the mean number of physician visits reported over the 25 year period. In the early 1950s the poor averaged 70 per cent as many visits as the nonpoor. From that time on their relative position improved; of major impetus were the federal financing initiatives during the mid-1960s. By 1970 the poor were averaging considerably more visits than the nonpoor (approximately one third). By 1976 this relative advantage appeared to decline somewhat.

The relative position of the poor in terms of the proportion who have seen a dentist during the year improved substantially during the 1950s and 1960s. However, less improvement is noted during the 1970s and their rate of use remains considerably below that of the nonpoor (about 60 per cent of parity).

In summary, Table 2 reinforces our conclusions from earlier analyses of the CHAS surveys that access of the poor relative to the nonpoor has improved significantly since the early 1950s. Substantial inequities persist, however, in the areas of dental services, prenatal care, and adequate third-party coverage; there have been no signs of these gaps narrowing in recent years. Further, dramatic increases observed in the 1950s and 1960s with respect to rate of hospital and physician use by the poor compared with the nonpoor did not appear to continue into the 1970s.

Table 3 (from NCHS) provides even more recent data, including 1978 and 1979, to illuminate whether these trends are continuing. Also, the use of a second major data source is important for confirmation because of the complexity of the access concept and the rudimentary state-of-the-art concerning its measurement.

The data in Table 3 tend to confirm the broad profile of equity on the part of the poor and nonpoor observed in Table 2. The poor continue to have a considerable dis-

TABLE 3. Selected Indicators of Access to Medical Care of the Poor* Compared With the Nonpoor: From the NCHS National Surveys† (1964, 1973, 1976, 1978, 1979)

Access Indicators	Ratio Score: Poor/Nonpoor‡				
	1964	1973	1976	1978	1979
Hospital discharges per 100 persons per year	1.10	1.52	1.47	1.51	—
Per cent visiting a physician in past two years	0.88	0.96	0.97	0.99	—
Physician visits per person per year	0.94	1.14	1.17	1.21	1.22
Per cent visiting a dentist in past two years	0.57	0.67	0.70	0.70	—
Dentist visits per person per year	0.44	0.61	0.61	0.61	—

* Source: Unpublished data from the Health Interview Survey, National Center for Health Statistics.

† Definitions of poor based on total family income of under: (1964) $3,000; (1973) $6,000; (1976) $7,000; (1978) $7,000; (1979) $7,000.

‡ Score less than 1.00 indicates the poor have less access; score greater than 1.00 indicates the poor have greater access.

advantage compared with the nonpoor with respect to dental service use. The leveling off of gains on this indicator between 1963–64 and 1976 appears to be sustained through 1978.

The NCHS hospital discharge data show a pattern similar to the CHAS-NORC hospital admission rate estimates—dramatic gains for the poor in the mid to late 1960s with a leveling off in the 1970s.

The NCHS data—as do CHAS-NORC* data not reported here—suggest that the proportion of the poor and nonpoor contacting a doctor at least once is approaching parity. The NCHS mean physician visit data similarly indicate there has been a sustained increase in the relative position of the poor through the 1960s; in the 1970s the rate of this increase began to decline to some extent.

To interpret more adequately the equity of the differences observed between the low and high income groups, controls for the variant levels of illness experienced by the respective groups and their differential age composition should be attempted. For example, the poor have traditionally reported more days of disability and higher rates of chronic illness than the nonpoor.[42] We have observed from recent overall mean physician visit and hospital use data that the poor now, in fact, use more services on the average than do the nonpoor. This appears to be a move in the right direction to achieve equity. Tables 4–6 present several alternatives to estimating equity between the high and low income groups, computing physician use rates adjusted for need and age. These indicators represent an effort to refine empirically and apply the concept of equity based on need that was discussed earlier.

Table 4 summarizes the rates of use relative to need for the poor and nonpoor from the CHAS surveys in contrast to physician-determined norms of "appropriate" access. The symptoms-response ratio reflects whether individuals in different age groups see a doctor more or less often for symptoms experienced than a panel of physicians thought was warranted.[19] A zero

---

* The CHAS-NORC and NCHS-HIS estimates of physician visits differ slightly in that the former excludes phone calls, while the latter includes them.

TABLE 4. Symptoms-Response Ratio for the Poor* Compared With the Nonpoor: From the CHAS National Surveys†
(1963, 1970, 1976)

Family Income Level	Symptoms-Response Ratio Score‡ 1963	1970	1976
Poor	−28.00	−14.00	−5.00
Nonpoor	−12.00	−11.00	−6.00

* Source: Reference 19.

† Definitions of poor based on total family income of under: (1963) $4,000; (1970) $6,000; (1976) $8,000. In each year approximately one-fourth of the population was defined as poor.

‡ A minus score indicates group saw a doctor less often for symptoms experienced than panel of physicians judged appropriate. A plus score indicates group saw a doctor more often than panel judged appropriate.

score on the index means individuals saw a doctor as often as the doctors thought necessary. A negative score means doctors were seen less often and a positive score indicates doctors were seen more often than doctors thought appropriate.

Table 4 provides a very encouraging view of the access gains of the poor between 1963 and 1976. The tendency for low-income individuals to see a doctor much less often (28 per cent less often in 1963, for example) than the physicians thought necessary has changed dramatically over time. According to the symptoms-response ratio for 1976, the poor and nonpoor had very similar profiles of appropriate contact for symptoms—5 and 6 per cent less often than the physicians judged appropriate, respectively. However, there is still a tendency for both low and high income individuals to see a physician less often for symptoms than clinically-based norms suggest is appropriate.

Tables 5 and 6 provide the NCHS alternative need-adjusted physician visit measures for years 1964 through 1979. Table 5 is a modified use-disability index that summarizes the ratio of mean physician visits to mean disability days for the respective age and income groups.† A higher score on the ratio means that the group had more visits relative to days of disability than those with a lower score. This ratio may be considered a physician-use indicator adjusted primarily for acute illness because the majority of the disability days reported are associated with acute conditions. In contrast, the basis for the data reported in Table 6 is mean physician visit estimates for the respective groups adjusted for the proportion with chronic illness in each of the groups.‡

According to Table 5, the poor have not achieved parity with the nonpoor in terms of a very general index of visits relative to reported days in bed or of limited activity. Poor children and young adults appear to have come closest to parity. This index shows considerable fluctuation rather than consistent improvement for the poor from the early 1960s through the late 1970s. Indeed, the ratios from the latest year (1979) are very similar to the initial ones shown for 1964.

According to the index adjusted for chronic illness in Table 6, there appears to have been substantial improvement in the status of the poor of all age groups since the early 1960s. The relative improvement for poor children—from 57 per cent of parity in 1964 to virtual parity (0.96) in 1979—is particularly noteworthy. As all of the poor groups approached or exceeded parity in the mid 1970s we observe fluctuations in the trends, but these fluctuations should be interpreted with caution. There is some suggestion that the gains by the poor in certain age groups cannot be maintained.

In sum, the three indices of physician visits adjusted in various ways for illness

---

† This modified ratio differs from the original in that it is not limited to individuals who had at least one day of disability in the year. See Reference 19, pp. 185-192, for a fuller description of how the original index was constructed.

‡ These adjustments are explained in *Health: United States, 1980.*[35]

TABLE 5. Physician Visits/100 Disability Days for the Poor* Compared With the Nonpoor, By Age Group: From the NCHS National Surveys†
(1964, 1973, 1976, 1978, 1979)

Age Group	Ratio Score:Poor/Nonpoor‡				
	1964	1973	1976	1978	1979
Under 17 years	0.78	0.72	0.79	0.77	0.78
17–44	0.69	0.70	0.72	0.74	0.74
45–64	0.49	0.47	0.55	0.60	0.46
65 and over	0.62	0.74	0.72	0.58	0.61

* Source: Unpublished data from the Health Interview Survey, National Center for Health Statistics.

† Definitions of poor based on total family income of under: (1964) $3,000; (1973) $6,000; (1976) $7,000; (1978) $7,000; (1979) $7,000.

‡ Score less than 1.00 indicates the poor have fewer visits per 100 disability days than nonpoor. Score greater than 1.00 indicates the poor have more visits per 100 disability days than the nonpoor.

(Tables 4–6) provide a somewhat mixed picture. The symptoms-response ratio (Table 4) and visits adjusted for chronic illness (Table 6) shows the poor making consistent progress from the early 1960s through the mid 1970s. Information from the later 1970s on the chronic illness measure fails to substantiate systematic and sustained gains for the poor in all age groups. It might be argued, according to the need concept of equity, that the higher scores for the poor are not justified. What would be disturbing, however, is the possibility that we are documenting the beginnings of a reversal of the earlier favorable trends; this could result in the poor systematically losing ground and regressing to their traditional disadvantaged position. The use-disability ratio (Table 5), like the chronic illness adjustment, fails to show systematic gains for the poor in the late 1970s but also does not support the view that the poor made systematic gains in the 1960s and early 1970s.

Tables 7–10 provide comparisons among 12 different communities and the national average for selected access indicators. These data are based on access surveys conducted in representative communities in which there was a community hospital-sponsored group practice funded by The Robert Wood Johnson Foundation

TABLE 6. Poor to Nonpoor Physician Visit Ratios* Adjusted for Chronic Illness by Age Group: From the NCHS National Surveys†:
(1964, 1973, 1976, 1978, 1979)

Age Group	Ratio Score Poor/Nonpoor‡				
	1964	1973	1976	1978	1979
Under 17 years	0.57	0.88	0.89	0.93	0.96
17–44	0.80	1.06	1.06	1.21	1.20
45–64	0.80	0.90	0.94	1.16	0.97
65 and over	0.75	0.87	1.01	0.85	0.90

* Source: Unpublished data from the Health Interview Survey; References 43, 44.

† Definitions of poor based on total family income of under: (1964) $3,000; (1973) $6,000; (1976) $7,000; (1978) $7,000; (1979) $7,000.

‡ Score less than 1.00 indicates the poor have fewer visits adjusted for their rate of chronic illness than the nonpoor. Score greater than 1.00 indicates the poor have more visits than the nonpoor.

TABLE 7. Proportion of Persons With a Regular Source of Care in Selected Communities (1979) Compared With the U.S. Average (1976) by Age Group*

	Community Characteristics			Ratio Score: Community/Nation[∥]			
				Age			
Community	Income Level[†]	Residence[‡]	Region[§]	Under 17 years	18–34	35–64	65 and over
A	Low	Metro	NE	1.02	1.08	0.92	0.99
B	Low	Metro	S	1.05	1.07	1.01	1.02
C	Low	Metro	W	0.79	0.71	0.83	0.98
D	Low	Nonmetro	NE	0.99	0.96	0.90	1.04
E	High	Metro	NC	1.05	1.04	1.00	0.95
F	High	Metro	NC	1.00	1.02	0.99	0.93
G	High	Metro	W	0.98	1.02	1.05	0.93
H	High	Nonmetro	W	0.84	0.75	0.81	0.90
I	High	Nonmetro	S	1.04	1.08	1.01	1.03
J	High	Nonmetro	NE	1.06	0.96	0.95	0.95

* Source: Unpublished preliminary data from Community Hospital Program Access Study, CHAS, University of Chicago; Reference 19.

[†] Low indicates less than $10,000; high indicates $10,000 or more.

[‡] Metro indicates community is within a Standard Metropolitan Statistical Area. Nonmetro indicates community is not within a Standard Metropolitan Statistical Area.

[§] Refers to major census divisions: NE = Northeast, NC = North Central, S = South, W = West.

[∥] Score less than 1.00 indicates proportion in community with a regular source is less than proportion in the country as a whole. Score greater than 1.00 indicates proportion in community is greater than in country as a whole.

as a part of its Community Hospital Program (CHP).[36,37]

These data were for the most part collected in late 1978 and early 1979.[§] The purpose of presenting these findings is to try to discover if there are continuing pockets of inequity among certain groups or types of communities that may not be apparent from the aggregate national statistics just reviewed. The communities in the Community Hospital Program represent a range of community types—high and low income, metropolitan (densely populated) and nonmetropolitan (sparsely populated), regionally diverse, etc.

Tables 7–9 show the ratios of the scores for the community compared with the national averages on the access measure of interest. A value of one means conformity with the national average, a value of less than one means the community score is lower, and a value of more than one, the community score is higher than the national average. The scores for various age groups in the respective communities are displayed in each table.

Assume that a score of plus or minus 10 per cent of the national average (range of 1.1–0.9) indicates that the community and the nation as a whole conform relatively closely. Most age groups within each of the CHP communities are near the national average in terms of the proportion of residents who report having a regular source of care (Table 7). Two communities from the Western region (C and H) are exceptions, with all age groups except the elderly showing limited access to a regular source of care. Young adults, 18–34 especially, were less likely to report having a regular place to go for medical care than was generally true nationally. In contrast, the elderly in these communities were within the plus or minus 10 per cent band. We may

---

[§] These data are preliminary and subject to change.

TABLE 8. Proportion of Persons With Third-Party Coverage of Medical Expenses in Selected Communities (1979) Compared With the U.S. Average (1976) by Age Group*

	Community Characteristics			Ratio Score: Community/Nation‖		
					Age	
Community	Income Level†	Residence‡	Region§	Under 17 years	18–34	35–64
A	Low	Metro	NE	1.04	1.05	0.95
B	Low	Metro	S	1.03	0.96	0.99
C	Low	Metro	W	0.81	0.74	0.88
D	Low	Nonmetro	NE	0.93	0.85	0.98
E	High	Metro	NC	1.12	1.12	1.10
F	High	Metro	NC	1.05	1.06	1.04
G	High	Metro	W	1.07	1.04	1.01
H	High	Nonmetro	W	0.94	0.90	0.99
I	High	Nonmetro	S	1.11	1.07	1.09
J	High	Nonmetro	NE	1.09	1.09	1.05

* Source: Unpublished preliminary data from Community Hospital Program Access Study, CHAS, University of Chicago; Reference 19.

† Low indicates less than $10,000; high indicates $10,000 or more.

‡ Metro indicates community is within a Standard Metropolitan Statistical Area. Nonmetro indicates community is not within a Standard Metropolitan Statistical Area.

§ Refers to major census divisions: NE = Northeast, NC = North Central, S = South, W = West.

‖ Scores less than 1.00 indicate proportion in community with third-party coverage is less than proportion in country as a whole. Scores greater than 1.00 indicate proportion in community is greater than in country as a whole.

be observing here the equalizing impact of Medicare at the community level.

The proportion of people with some form of public or private third-party coverage was also lower in Community C—a low income inner city area—than was the case nationally (Table 8). Among the other communities there was greater correspondence in the rates of coverage to the national average. However, it appears that the higher income communities tend to have a higher proportion of their residents reporting coverage than the low income communities.

The proportion of community residents who had seen a doctor within the year was within 10 per cent of the national average for the majority of age groups in all communities (Table 9). The major exception to this trend was the generally high proportion of children in the high income communities seeing a doctor. On the "deficit side," adults (except the elderly), particularly in the low income communities, tend to have contact rates below the national average. The most extreme case is the adult group in Community A (a low income inner city area in the Northeast) with a contact rate that is 86 per cent of the national average.

Table 10 compares the symptoms-response index in each community with the norms of appropriate response provided by the physician panel. It shows that there is considerable variance by community and age group in the symptoms-response ratio scores.‖ Some particular subgroups are considerably below this figure, e.g., 35–64-year-olds in Community D (a low income relatively rural area in the Northeast), and some are considerably above it, e.g., 18–34-year-olds in Community A. Despite unique community differences, some patterns reflective of national

---

‖ This is a revised version of the ratio shown in Table 4 that is based on a reduced number of the symptoms included in the national study.

TABLE 9. Proportion of Persons Visiting an M.D. Within a Year in Selected Communities (1979) Compared With the U.S. Average (1976) by Age Group*

	Community Characteristics			Ratio Score: Community/Nation‖			
				Age			
Community	Income Level†	Residence‡	Region§	Under 17 years	18–34	35–64	65 and over
A	Low	Metro	NE	1.02	0.91	0.86	1.06
B	Low	Metro	S	1.03	1.02	1.04	1.09
C	Low	Metro	W	0.99	0.87	0.96	1.01
D	Low	Nonmetro	NE	1.05	1.00	0.91	0.97
E	High	Metro	NC	1.17	0.94	1.01	0.98
F	High	Metro	NC	1.06	0.99	0.93	0.94
G	High	Metro	W	1.24	1.01	1.03	0.89
H	High	Nonmetro	W	1.15	1.02	0.98	0.94
I	High	Nonmetro	S	1.18	1.04	0.99	1.02
J	High	Nonmetro	NE	1.17	1.03	1.04	1.10

* Source: Unpublished preliminary data from Community Hospital Program Access Study, CHAS, University of Chicago; Reference 19.

† Low indicates less than $10,000; high indicates $10,000 or more.

‡ Metro indicates community is within a Standard Metropolitan Statistical Area. Nonmetro indicates community is not within a Standard Metropolitan Statistical Area.

§ Refers to major census divisions: NE= Northeast; NC = North Central, S = South, W = West.

‖ Score less than 1.00 indicates proportion in community visiting a physician is less than proportion in country as a whole. Score greater than 1.00 indicates proportion in community is greater than proportion in country as a whole.

findings do emerge[19]: higher scores among children and those in metropolitan areas. and lower scores for adults and individuals in nonmetropolitan areas.

In summary, Tables 2–10 suggest that according to the need and similar-treatment-for-similar cases criteria of equity, there have indeed been considerable improvements in the equity of access profile in the U.S. over the past 25 years. These caveats are also borne out by these data:

1. The poor are much less apt than the nonpoor to see a dentist during a year, (for pregnant women) to visit a doctor during the first trimester of pregnancy, or to have adequate private or public third-party coverage.

2. The relative gains of the poor in terms of overall physician visit rates may be leveling off or, to some extent, diminishing when compared with the nonpoor.

3. Persons with low-incomes may be using services in response to symptoms about as appropriately as persons with high-incomes, but other indicators suggest that relative to experienced disability or chronic illness, complete parity between income groups may still not have been achieved for all age groups.

4. Pockets of inequity continue to exist, depending on the measure of access that is considered, the type of community in which people live, and the social and demographic characteristics of persons within the community.

Governmental Programs and Equity

Clearly the two major government financing initiatives begun in 1966 that have most influenced the population's use of medical services are Medicare for the elderly (and later the disabled) and Medicaid for those with low incomes. In the later 1970s and 1980, reports of increasing state restrictions on eligibility, benefits, and reimbursements for the Medicaid Program prompt us to examine summary trends in Medicaid and Medicare performance that might change the overall access picture in the United States. Also, changes

TABLE 10. Symptoms-Response Ratio for Selected Communities (1979) by Age Group*

	Community Characteristics			Symptoms-Response Ratio Score‖			
				Age			
Community	Income Level†	Residence‡	Region§	Under 17 years	18–34	35–64	65 and over
A	Low	Metro	NE	26	34	−3	28
B	Low	Metro	S	18	−12	−1	24
C	Low	Metro	W	4	−11	11	8
D	Low	Nonmetro	NE	1	−15	−24	−17
E	High	Metro	NC	18	13	−17	10
F	High	Metro	NC	9	−7	−1	−12
G	High	Metro	W	11	−17	−10	−23
H	High	Nonmetro	W	−12	−22	−14	1
I	High	Nonmetro	S	6	7	−17	6
J	High	Nonmetro	NE	14	−8	−9	4
National Average				−1	−11	−6	0

* Source: Unpublished preliminary data from Community Hospital Program Access Study, CHAS, University of Chicago; Reference 19.

† Low indicates less than $10,000; high indicates $10,000 or more.

‡ Metro indicates community is within a Standard Metropolitan Statistical Area. Nonmetro indicates community is not within a Standard Metropolitan Statistical Area.

§ Refers to major census divisions: NE = Northeast, NC = North Central, S = South, W = West.

‖ A minus score indicates community residents saw a doctor less often for symptoms experienced than a panel of physicians judged appropriate. A plus score indicates community residents saw a doctor more often than panel judged appropriate.

in various other governmental efforts to provide service to particularly vulnerable groups and those who "fall through the cracks" between private and government financing systems (neighborhood and later community health centers, maternal and child health care, migrant and rural health care) have the potential to influence the access picture. Consequently, we examined trends in the Community Health Centers Program in particular, as well as Medicaid and Medicare.

Table 11 shows steady increases in Medicaid payments from about $3.5 billion in 1968 to over $20 billion in 1979. The number of Medicaid recipients more than doubled from 11.5 million in 1968 to about 23.5 million in 1976. However, in each of the last three years for which we have information the number declined, dropping to a little more than 21 million in 1979. Although we were cautioned that the apparent recent drop in recipients may be influenced by modified reporting in some states,[45] the impact of these changes on the overall summary figures is not clear. Further, the trend the past few years is parallel to the reports in numerous states of fiscal problems in the Medicaid Program with resultant cuts in services.

The annual payments per recipient have grown each year from $300 in 1968 to $950 in 1979. When these payments are adjusted for the increasing price of medical care the increase is not consistent. The constant dollar value of services received in 1979 is actually $424 or about 40 per cent more than in 1968.

Since its inception, Medicare has increased its enrollment each year starting with 19 million in 1966 and reaching almost 27 million in 1978 (Table 12). Contributing to the increase has been the growing number of persons 65 years of age and

TABLE 11. Medicaid Payments Adjusted for Increases in Recipients and Prices, Fiscal Years 1968–1978

Fiscal Year	Total (Federal and State) Medicaid Payments (Thousands)*	Yearly Number of Medicaid Recipients (Thousands)†‡	Medical Care Price Index§	Per Cent Growth of Medical Care Costs	Annual Payments Per Medicaid Recipient ($)	Payment Per Recipient in Constant Dollars
1968[‖]	3,451,376	11,500	100.0	6.1	300	300
1969[‖]	4,351,486	12,060	106.9	6.9	331	338
1970[‖]	5,093,901	14,507	113.7	6.4	351	309
1971	6,345,199	17,965	121.0	6.4	353	292
1972	7,346,131	17,990	124.9	3.2	408	327
1973	8,713,761	18,818	129.8	3.9	463	357
1974	9,737,398	20,842	141.9	9.3	467	329
1975	12,086,166	21,197	158.9	12.0	570	359
1976	13,977,348	23,462	174.1	9.6	596	342
1977	16,354,599	22,814	190.8	9.6	717	376
1978	17,965,000	22,198	206.4	8.9	809	392
1979	20,200,000	21,200	224.5	8.7	953	424

Source: Reference 45; Personal Communication: U.S. Department of Health and Human Services, Health Care Financing Agency, Medicaid Bureau, Region V (Chicago, Illinois).

* Source: "State Expenditures for the Medical Assistance Program" except for FY 1978 which comes from the Appendix to the Budget of the U.S. Government for FY 1980. See Technical Notes 1, 2 and 5 in Appendix.

† Includes some recipients of aid under nonfederally matched programs.

‡ Source: "Medicaid State Tables" except for 1971, 1977 and 1978 which are from "Medicaid Statistics." See Technical Notes 3, 4, 5, 6 and 7 in the Appendix.

§ Bureau of Labor Statistics Medical Care Price Index with adjustments to make 1968 = 100.

[‖] Includes payments under the Kerr-Mills program.

over as well as the inclusion in the program of other categories of people (those on kidney dialysis and the disabled).

Another Medicare-based indicator of access is the proportion of enrollees who chose to purchase supplementary medical insurance. This proportion has increased each year from 92.4 per cent in 1966 to 97.4 per cent in 1978. The average monthly premium for this supplementary insurance rose from $3.00 in 1966 to about $8.00 in 1978. Adjusting by the Consumer Price Index for all items suggests that recipients in 1978 were paying about one-third more for their medical insurance in 1978 than was the case in 1966.

Total reimbursements for hospital and supplemental medical insurance reached almost $24 billion in 1978. The average reimbursement per enrollee increased from $268 in 1968 (the first full year experience shown in Table 12) to almost $900 in 1978. Adjustment for inflation in the price of medical care suggests that the amount spent per recipient per year in 1966 dollars actually increased from $236 in 1968 to $381 in 1978, an increase of around 60 per cent.

The last indicator of Medicare access we examine is the deductible for hospital services, or the amount the hospitalized patient must pay before Medicare takes over. It was $40 in 1966 and had increased to $144 by 1978. Adjusting for inflation, the increase was from $40 to $72.

The Community Health Centers Program, although much smaller than the overall Medicaid Program in terms of expenditures and users, showed steady increases in the number of people its various projects served—from 1.2 million in 1974 to 4.2 million in 1980 (Table 13). Its distribution of revenues suggests that many of its services and the patients it serves are not covered under Medicare and Medicaid; in 1980 about 16 per cent of its revenues came from Medicare and Medicaid. Despite the increases in users (250 per cent) and encounters (200 per cent) between 1974 and 1980, appropriations in constant 1974 dollars were actually less in 1980 (174 million) than in 1974 (197 million).

Thus, our examination of Medicare and the Community Health Centers Program suggests that coverage in terms of the numbers of people served and services per recipient continued to increase in recent years. Potential access barriers in Medicare might be reflected, however, in the increasing cost of deductibles for hospitalization and the supplementary medical insurance premiums. In the case of the Community Health Centers Program, appropriations do not appear to be keeping pace with the number of services provided, although the Centers are securing revenues from a number of different sources.

The Medicaid picture is different. Although the payment per recipient in constant dollars showed some increase in recent years (about 24 per cent between 1976 and 1979), the number of recipients apparently declined by over two million in this same period. Assuming this decline is not simply a matter of record-keeping, it may represent a trend to increasing access barriers with time. During a similar time period (1976–1978) the number of people with below poverty level incomes declined by only one-half million.[50] Given relatively high rates of unemployment, a stagnant economy, and difficulties of local governments in maintaining their own medical care programs, it seems unlikely that other sources can cope with the decline in service delivery and financing resulting from Medicaid cutbacks.

Health Policy and the Current Profile of Equity

Issues to be addressed in any policy deliberations concerning the problems of

TABLE 12. Trends in the Medicare Program

Year	Number Enrolled for Hospital Insurance	Per Cent of Hospital Enrolled with Supplemental Medical Insurance	Total Reimbursement for Hospital and Supplemental Medical Insurance	Average Expenditure/ Recipient (Current $)	Average Expenditure/ Recipient (1966 $)	Inpatient Hospital Deductible (Current $)	Inpatient Hospital Deductible (1966 $)	Supplementary Medical Insurance Average Premium/Month (Current $)	Supplementary Medical Insurance Average Premium/Month (1966 $)
1966*	19,082	92.4	886,943	46	46	40	40	3.00	3.00
1968	19,770	95.1	5,289,026	268	236	40	37	4.00	3.73
1970	20,361	96.2	6,605,683	324	251	52	43	4.70	3.93
1972	21,115	96.4	8,165,978	387	273	68	53	5.50	4.27
1973†	23,302	96.5	8,642,782	371	252	68	50	6.05	4.70
1975	24,640	97.0	14,523,573	589	326	92	59	6.70	4.04
1976	25,312	97.2	17,350,766	685	346	104	59	6.95	3.97
1977	26,094	97.2	20,869,055	800	369	124	66	7.45	3.99
1978	26,777	97.4	23,955,521	895	381	144	72	7.95	4.00

Source: References 44, 46–50. * Program in effect half the year. † From 1973 on statistics include disabled persons under 65 years of age.

TABLE 13. Community Health Centers

	1974	1976	1978	1980
Appropriations (Current $)	197m	197m	247m	320m
Appropriations (1974 $)	197m	155m	162m	174m
Users	1.2m	2.2m	3.0m	4.2m
Encounters	5.6m	10.0m	13.3m	16.8m
Projects	157	302	591	872

1976 Survey of CHC Users Showed:
71% under poverty level
80% minority
49% have no employed family member
41% under 18

CHC revenues from first and third parties

	1974	1977	1978	1979	1980
Medicare		2.9%	2.3%	2.4%	2.5%
Medicaid		15.1	12.4	12.7	13.3
Title XX		0.3	0.3	0.3	0.3
Private insurance		6.1	6.5	6.7	7.0
Other third party		2.7	3.2	3.3	3.4
Patient fees		9.3	12.5	12.9	13.5
Total (%)	17.4	36.4	37.2	38.3	40.0

Note: In 1974 NCHS began a major management push to improve project collections from first and third parties, and later years show results of these efforts. Clearly, however, Medicaid is *not* providing the major financing of CHCs. In fact, the Medicaid dropoff after 1977 reflects: 1) increased CHC activity in rural areas with lower percentages of Medicaid recipients, and 2) increasing state restrictions on eligibility, benefits, and reimbursements.

Source: Personal Communication: Ronald H. Carlson, Associate Administrator for Planning, Evaluation and Legislation, Health Services Administration, Department of Health and Human Services.

equity include: 1) are the problems of sufficient magnitude or importance to warrant the expenditure of resources, and 2) if so, what means might be effective in solving them?

Consideration of the first issue must draw upon some of the questions raised in the initial sections of this paper. Inequities do remain that are based upon the need and similar treatment criteria. The magnitude of the problem is less than in the past, but there is no evidence that substantial narrowing of existing gaps is taking place or that current trends will continue should there be significant reductions in programs previously responsible for initiating and sustaining these trends.

Practical questions to consider in deciding whether the problems that remain are important enough to warrant further investment of resources include:

1. What are the value judgments implied in defining a particular observed "inequity" as a problem?
2. What criteria should be used to measure how much of a difference signals a "serious" problem, i.e., plus or minus "X" per cent of the national average?
3. Will addressing this problem bring about other positive outcomes, i.e., improved health status or more consumer satisfaction?
4. Are the resources to bring about the desired change available either from the individual or the delivery system?

5. Will achieving the objective of equity of access mean that other (e.g., cost or quality) goods are compromised?

6. What are the likely outcomes if no attention is directed to the problem or if existing programs dealing with the problem are discontinued?

These are questions of practical importance in fluid and stressful economic and political times. Assuming that some of the issues identified here are ones that are singled out for attention, we are then called upon to address the second question of the means by which those desired ends can be effected. As mentioned in preceding sections of the paper, the contribution that health services research has made (more or less successfully) in answering this question are efforts to model the "if-then" of various policy choices. There is evidence, for example, that financing initiatives have been responsible primarily for narrowing the overall use differentials by income. On the other hand, health care reorganization strategies such as community health centers or comprehensive family-centered HMOs also may have contributed to these outcomes. Efforts to focus on solutions to the problems that remain should build on existing empirical work that attempts to estimate the coefficient of association, and/or causality, between certain desired outcomes and mutable policy-relevant means to effect them.

Future empirical work on the equity of access concept could aid the process of decision-making just outlined by 1) refining further the means of measuring need and use, and the relationship between the two; 2) developing and applying various criteria of appropriateness; 3) estimating program impact, and 4) developing operational indicators of explicit tradeoffs in the cost, quality, and access objectives.

Forces in our society that portend increasing emphasis on cost control at the expense of quality and, in particular, access objectives include: 1) the current conservative political climate highlighted by the efforts to cut government spending; 2) the concern about inflation, with medical care singled out as a major contributor; 3) increasing life expectancy and reduced infant mortality and disease-specific death rates (such as heart disease); and 4) the growing emphasis on good health habits (diet, exercise, self diagnosis and treatment).

Under these circumstances even the concept of medical care as a right may be increasingly questioned. The Reagan administration has proposed substantial cuts in Medicaid spending. The criteria of social justice dominant in our society with respect to the distribution of health services may shift toward the marketplace, merit, and societal contribution, and therefore away from need and similar treatment.

The shift might be buttressed by the evidence that the poor currently have higher admission rates and more physician visits than the rest of the population; an inability to show that much medical care has a positive impact on traditional health status measures; continued improvement in these health status measures (mortality rates); and the increasingly held belief that the individual may be largely responsible for good health status through personal habits or "meritorious actions."

Our findings suggest that the access picture is much more complex than the simple gross hospital and physican use rates indicate. Further, we may be at a critical point as far as slowing or reversing trends in access that have been documented over the past 30 years. The indicators in this report (and it is hoped that better ones will be developed) need to be continually updated and analyzed. This activity will not necessarily alter the criteria currently in vogue for judging equity of access and resulting policy and programs. At least these indicators can relay in a timely fashion to our leaders and the public some of the costs (or benefits) of policy decisions concerning health services delivery and financing.

Acknowledgments

The authors acknowledge the contributions of Ron Wilson (National Center for Health Statistics), Ron Carlson and Ann Zuvekas (Public Health Service) who provided data reported in this paper; of Charles Harris (Methodist Medical Center, St. Louis, MO) for sharing his work on ethics and health policy with us; of Susan Turk and Terry Herold (CHAS) for assembling references and information for the paper; and of Joyce Van Grondelle (CHAS) for typing the manuscript.

References

1. Blackstone WT. On health care as a legal right: an exploration of legal and moral grounds. Georgia Law Review 1976;10:391.

2. Sidel VW. The right to health care: an international perspective. In: Bioethics and Human Rights. Boston: Little, Brown, 1978.

3. White RL. Right to Health: The Evolution of an Idea. Iowa City: The University of Iowa, 1971.

4. Carlson R. The End of Medicine. New York: John Wiley and Sons, 1975.

5. Illich I. Medical Nemesis: The Expropriation of Health. Toronto: Bentam Books, 1976.

6. Blumstein JF, Zubkoff M. Public choice in health: problems, politics and perspectives on formulating national health policy. J Health Polit Policy Law 1979;4:382.

7. Mechanic D. Rationing health care: public policy and the medical marketplace. Hastings Center Report 1976;6:34.

8. Sparer EV. The legal right to health care: public policy and equal access. Hastings Cent Rep 1976;6:39.

9. McGarvey MR. Some considerations regarding ethics and the right to health care. In: Bioethics and Human Rights. Boston: Little, Brown, 1978.

10. Gustafson JM. Interdependence, finitude, and sin: reflections on scarcity. Journal of Religion 1977;57:156.

11. Hiatt HH. Protecting the medical commons: who is responsible? N Eng J Med 1975;293:235.

12. Mechanic D. Ethics, justice, and medical care systems. Ann *AAPSS* 1978;437:74.

13. Weinstein MC, Stason WB. Allocating resources: the case of hypertension. Hastings Cent Rep 1977;7:24.

14. Levine C. Ethics and health cost containment. Hastings Cent Rep 1979;9:10.

15. Navarro V. Health and the corporate society. Social Policy 1975;5:41.

16. Pellegrino ED. Medical morality and medical economics. Hastings Cent Rep 1978;8:8.

17. Aday LA, Andersen R. A framework for the study of access to medical care. Health Serv Res 1974;9:208.

18. Aday LA, Andersen R. Development of Indices of Access to Medical Care. Ann Arbor MI: Health Administration Press, 1975.

19. Aday LA, et al. Health Care in the U.S.: Equitable for Whom? Beverly Hills CA: Sage Publications, 1980.

20. Outka G. Social justice and equal access to health care. Perspec Biol Med 1975;18:185.

21. Navarro V. The political economy of medical care: an explanation of the composition, nature and functions of present health sector of the United States. Int J Health Serv 1975;5:65.

22. Navarro V. Justice, social policy, and the public's health. Med Care 1977;15:363.

23. Fuchs V. Concepts of health—an economist's perspective. J Med Philos 1976;1:229.

24. Garfinkel I. Equal access, minimum provision, and efficiency in financing medical care. J Hum Resour 1972;7:242.

25. Fried C. Equality and rights in medical care. Hastings Cent Rep 1976;6:29.

26. Fein R. On achieving access and equity in health care. Milbank Memorial Fund Quarterly 1972;50:157.

27. Bryant JH. Principles of justice as a basis for conceptualizing a health care system. Int J Health Serv 1977;7:707.

28. Harris C. Justice and Value in Health Policy. Columbia: University of Missouri, 1980.

29. Verbrugge LM. Equity of Access to Health Care: A Research Review and Prospectus. Ann Arbor: Institute for Social Research, University of Michigan, 1980.

30. Aguirre MN, et al. Response: The Consumer Viewpoint: "What is Health Care and What do we Want?" DHEW Publication No. (HRA) 79-632. Washington, DC: U.S. Government Printing Office, 1980.

31. Mechanic D. Correlates of physician utilization: why do major multivariate studies of physician utilization find trivial psychosocial and organization effects? J Health Soc Behav 1979;20:387.

32. Jonsen AR, Butler LH. Public ethics and policy making. Hastings Cent Rep 1975;5:19.

33. Sloan FA. Toward measures of equity in the delivery of health care services. In: Nourishing the Humanistic in Medicine. Pittsburgh: University of Pittsburgh Press, 1979.

34. Andersen R, et al. Two Decades of Health Services: Social Survey Trends in Use and Expenditure. Cambridge, MA: Ballinger, 1976.

35. National Center for Health Statistics. Health: United States, 1980. DHHS Publication No. (PHS) 81-1232. Washington, DC: U. S. Government Printing Office, 1980.

36. Aday LA, et al. Overview of a design to evaluate the impact of community hospital-sponsored primary care group practices. Medical Group Management 1978;256:42.

37. Block JA, et al. Physicians and hospitals: providing primary care. Medical Group Management 1978;25:35.

38. Andersen R, Anderson OW. A Decade of Health Services: Social Survey Trends in Use and Expenditure. Chicago: The University of Chicago Press, 1967.

39. Andersen R, Anderson OW. Trends in the use of health services. In: Handbook of Medical Sociology. Englewood Cliffs: Prentice-Hall, 1979.

40. National Center for Health Statistics. Persons hospitalized by number of hospital episodes and days in a year: United States—July 1965-June 1966. Series 10, No. 50. Washington, D.C.: U.S. Government Printing Office, 1969.

41. Phelps CE. Effects of insurance on demand for medical care. In: Equity in Health Services: Empirical Analyses in Social Policy. Cambridge, MA: Ballinger, 1975.

42. Newacheck PW, et al. Income and illness. Med Care 1980;18:1165.

43. National Center for Health Statistics. Health characteristics of persons with chronic activity limitations: United States—1974. Series 10, No. 112. Washington, DC: U.S. Government Printing Office, 1976.

44. National Center for Health Statistics. Health: United States, 1979. DHEW Publication No. (PHS) 80-1232. Washington, DC: U.S. Government Printing Office, 1980.

45. Health Care Financing Administration. Data on the Medicaid Program: Eligibility, Services and Expenditures. Baltimore: Medicaid/Medicare Management Institute, Health Care Financing Administration, 1979.

46. Social Security Administration. Social Security Bulletin, Annual Statistical Supplement, 1975. Washington, DC: U.S. Government Printing Office, 1975.

47. Social Security Administration. Social Security Bulletin, Annual Statistical Supplement, 1977–79. Washington, DC: U.S. Government Printing Office, 1979.

48. U.S. Department of Commerce. Statistical Abstract of the United States, 1978. Washington, DC: U.S. Government Printing Office, 1978.

49. Waldo DR. Health Care Financing Trends 1980;2. Washington, D.C.: Office of Research, Demonstrations and Statistics, HCFA.

50. U.S. Department of Commerce. Population Profile of the United States, 1979. Current Population Reports. Series P-20, No. 350. Washington, DC: U.S. Government Printing Office, 1980.

-/ C5170-1 -

Medical Care

January 1982
Vol. 20 No. 1
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

The Efficacy of Second-Opinion Consultation Programs: A Cost-Benefit Perspective

Impact of a Mandatory Second-Opinion Program on Medicaid Surgery Rates

The Effect of Concurrent Feedback in Reducing Inappropriate Hospital Utilization

Medical Information Systems: Assessing Impact in the Areas of Hypertension, Obesity and Renal Diseases

Assessing Medical Practices Through PSRO Cooperative Studies

Periodic Medical Review: The Impact in Pennsylvania

Alcoholism, Morbidity and Care-Seeking

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $40.00; special price U.S. students, $23.00; all other countries except Japan, one year, $50.00. Single copies $5.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

January 1982
Volume 20, Number 1

CONTENTS

EDITORIAL

Second-Opinion Programs: Beyond Cost-Benefit Analyses. *Robert H. Brook and Kathleen N. Lohr* 1

ORIGINAL ARTICLES

The Efficacy of Second-Opinion Consultation Programs: A Cost-Benefit Perspective. *Hirsch S. Ruchlin, Madelon Lubin Finkel and Eugene G. McCarthy* 3

Impact of a Mandatory Second-Opinion Program on Medicaid Surgery Rates. *Suzanne Grisez Martin, Michael Shwartz, Bernadette J. Whalen, Deborah D'Arpa, Greta M. Ljung, John Holden Thorne and Anne E. McKusick* 21

The Effect of Concurrent Feedback in Reducing Inappropriate Hospital Utilization. *Joseph D. Restuccia* 46

Medical Information Systems: Assessing Impact in the Areas of Hypertension, Obesity and Renal Disease. *James L. Rogers, Olga M. Haring, Paul M. Wortman, Royce A. Watson and John P. Goetz* 63

Assessing Medical Practices Through PSRO Cooperative Studies. An Evaluation of Cesarean Births in Nine PSRO Areas. *William F. Jessee, C. W. Nickerson and William S. Grant, for the PSRO Coordinating Group* 75

Periodic Medical Review: The Impact in Pennsylvania. *Michael A. Calogero and Daniel B. Hill* 85

Alcoholism, Morbidity and Care-Seeking. The Inpatient and Ambulatory Service Utilization and Associated Illness Experience of Alcoholics and Matched Controls in a Health Maintenance Organization. *Sandra L. Putnam* 97

LETTER TO THE EDITOR

On Assessing the Performance of Medical Care Systems. *James I. Hudson.* 122

BOOK REVIEWS

Toward Rational Technology in Medicine: Considerations for Health Policy, by H. David Banta, Clyde J. Behney and Jane Sisk Willems. *Eric B. Larson* 122

Assessing the Contributions of the Social Sciences to Health, edited by M. Harvey Brenner, Anne Mooney and Thomas J. Nagy. *Robert L. Eichhorn* 123

The Planning of Health Services: Studies in Eight European Countries, edited by G. McLachlan. *L. S. Rosenfeld* 124

Women, Work, and Health: Challenges to Corporate Policy, edited by Diana Chapman Walsh and Richard H. Egdahl. *Gail Gordon* 125

MEDICAL CARE
January 1982, Vol. XX, No. 1

Original Articles

The Efficacy of Second-Opinion Consultation Programs: A Cost–Benefit Perspective

HIRSCH S. RUCHLIN, PH.D.,* MADELON LUBIN FINKEL, PH.D.,† AND EUGENE G. MCCARTHY, M.D., M.P.H.‡

This study evaluates a mandatory second-opinion consultation program administered on behalf of a large Taft-Hartley welfare fund providing medical care coverage for 120,000 beneficiaries and covered dependents. During a two-year intake period (1977–1978), 2,284 individuals received second-opinion consultations for an elective surgical procedure recommended by a first-contact physician or surgeon. Of this group, 366 received a nonconfirmation of their need for surgery. Medical claims data were available for 342 individuals in this group, and they constitute the base for the current analysis. A comparable number of individuals who received a positive confirmation were randomly selected and served as a control for estimating program savings. Both groups were followed for a one-year period from the date of their consultations. Total program savings were estimated at $534,791. Of this amount, medical care utilization savings were $361,756 and productivity savings were $173,035. The cost of the program was $203,300, yielding a benefit–cost ratio of 2.63. These findings indicate that mandatory second-opinion consultation programs, which are consumer oriented and intervene before care is rendered, are clearly cost-effective.

SECOND-OPINION CONSULTATION PROGRAMS for elective surgery are designed primarily to assist patients in making a more informed decision once elective surgery has been recommended. The principal goal of such programs is to provide patients with additional information on their need for surgery and, where applicable, on alternate modes of treatment. In addition to improving quality of care, second-opinion consultation programs, by reducing the rate of elective surgery, can also lead to significant economic savings.

Studies reported in the literature indicate that approximately 30 per cent of the individuals who voluntarily seek a second-opinion consultation, and approximately 18 per cent of those who are re-

---

This research was supported by Contract No. 600-75-0175 from the Health Care Financing Administration, Department of Health and Human Services.

* Professor of Economics in Public Health, Department of Public Health, Cornell University Medical College, New York, New York.

† Assistant Professor, Department of Public Health, Cornell University Medical College, New York, New York.

‡ Clinical Professor, Department of Public Health, Cornell University Medical College, New York, New York.

Address for reprints: Hirsch S. Ruchlin, Ph.D., Department of Public Health, Cornell University Medical College, 1300 York Avenue, New York, NY 10021.

0025-7079/82/0100/0003/$01.40 © J. B. Lippincott Co.

quired by their insurance carriers to seek a second opinion, receive a nonconfirmation§ of their need for surgery.[1-7] The magnitude of the nonconfirmation rate is taken as prima facie evidence of the presence of "surplus surgery." Follow-up studies of individuals participating in the largest mandatory second-opinion consultation program—the one conducted by Cornell University Medical College in conjunction with several Taft-Hartley security welfare funds‖ in the greater New York City area—indicate that of those not confirmed for surgery, 61 per cent had not had the surgery performed one year from the date of their second-opinion consultation. Moreover, 51 per cent of those not confirmed for surgery reported that they neither had the surgery performed nor received any form of medical treatment since their second-opinion consultation. For those who voluntarily sought a second-opinion consultation, these statistics are 78 per cent and 41 per cent, respectively.[8] On the basis of these data, one can hypothesize that second-opinion programs should yield significant economic savings. The current study presents data supporting this hypothesis.

In evaluating a second-opinion consultation program, one must recognize the fundamental limitations inherent in analyzing medical or economic outcome when the election of a second-opinion consultation is voluntary. Experience indicates that less than 5 per cent of those eligible for participation in a voluntary second-opinion consultation program take advantage of this benefit. By the very nature of the program's design, individuals who voluntarily seek a second opinion are a self-select group. Evaluations based on the results of a mandatory program do not suffer from such inherent selection biases and consequently have greater validity and generalizability.

The current study is based on the experience of a large mandatory second-opinion consultation program administered by the Building Service Local 32B-J welfare fund. This fund provides health benefit coverage for 120,000 individuals (employees and covered dependents) living in the greater New York City metropolitan region. The methodology used in the current study is outlined in the next section. This is followed by a description of the study population and the data base, and analyses of program savings and costs. The findings emanating from these analyses are presented and discussed in the concluding sections.

Methodology

The primary focus of the economic evaluation is to ascertain the cost-containment potential of a mandatory second-opinion consultation program. Consequently, the following rule was adopted in the modeling of the program: Any recommendation emanating from the consultation process that could potentially lead to reduced health care expenditures is considered to be a nonconfirmation of the original recommendation for surgery, e.g., surgery performed on an ambulatory basis. This is the case even if the new recommendation results in a modified surgical regimen or if the initial recommendation for surgery is ultimately accepted after trying a recommended medical treatment.

As a result of receiving a second-opinion consultation, individuals are classified into one of two categories: confirmed or not-confirmed.# The evaluation focus is thus centered on ascertaining whether the *receipt* of a second-opinion consultation generates economic savings (benefits). These savings should materialize through the election of either less surgery, less expensive surgery or less expensive medical care.¶

---

§ A nonconfirmation results when the consultant surgeon feels that the initially recommended surgery is not the most appropriate therapeutic next step. The consultant may feel that medical treatment or an alternate surgical procedure is more appropriate, or that there is no pathological justification for surgery.

‖ These funds were created by the Taft-Hartley Act of 1948 and are the principal mechanisms by which private industry in the United States provides fringe benefit coverage to workers and their dependents. Management and labor are equally represented in each fund.

Within both the confirmed and the not-confirmed groups, one can delineate various subgroups based on the individual's actions following the receipt of the second-opinion consultation. Individuals who received a confirmation of their need for surgery could elect or reject surgery. Not-confirmed individuals can be placed in one of four categories: those who had no surgery, those who elected a modified surgical procedure, those who elected the originally recommended surgery after trying an intervening medical treatment and those who elected the originally recommended surgery without any intervening medical care. Although the cost–benefit evaluation reported here will be based solely on data pertaining to the entire not-confirmed and confirmed groups, medical care utilization data for each subgroup are of interest and are reported in the Appendix.

Resources saved by the avoidance of surgery or the election of a less expensive treatment constitute the economic measure of the program's success. These savings are ascertained by subtracting the total cost profiles generated by the not-confirmed group from the total cost generated by the confirmed individuals. If the program is successful, in an economic sense, the not-confirmed group will exhibit lower expenditure patterns than the confirmed group. Two major benefit categories were delineated for the current study: 1) savings of health-related resources and 2) avoidance of work loss and restricted-activity days. The first type of saving encompasses reduced surgical and medical care, hospitalization, drug and therapy use, and ancillary service use attributable to the second-opinion consultation program. Savings stemming from reduced work loss and restricted-activity days avoided reflect the indirect benefits gained. The cumulative monetary value of both types of benefits is compared with the cost of the second-opinion consultation program. Program cost consists of the administrative costs incurred to operate the program, the expenditures for the second-opinion consultation visit and any ancillary tests ordered as part of the consultation, travel costs incurred by program participants in obtaining the consultation and wage loss (i.e., opportunity costs) due to work time missed in order to obtain the second-opinion consultation.

Findings reported in this study identify utilizer and cohort patterns. Data presented for utilizers indicate only specifically *reported* services consumed and work loss and restricted-activity days incurred. Cohort data represent actual reported utilization and, where necessary, proxied values. Means reported for the cohorts were derived by summing all reported and proxied data and dividing this statistic by the total number of utilizers and *nonutilizers* in that group. These means indicate average expected utilization and cost spread over the entire group. Such a

---

Individuals participating in the second-opinion program may request an additional (i.e., third) consultation. In such cases, the opinion of the last (i.e., second) consultant determines whether the intial recommendation for surgery is considered to be confirmed.

¶ One should not expect every "not-confirmed" individual to decide not to have surgery. Similarly, not every "confirmed" individual will elect to have surgery. However, one expects that the majority of individuals in each group will follow the consultant's recommendation. For the current study population, 89 per cent of the confirmed individuals elected surgery, while 85.1 per cent of the not-confirmed group followed the recommendation of the second-opinion consultant. As the program being evaluated permits the individual to decide whether or not to elect surgery, both behavioral patterns must be recognized within the confirmed and not-confirmed groups.

statistic has relevance for actuarial and insurance analyses. Means reported for the utilizers depict average utilization and cost incurred only by those individuals actually reporting data. As both the actuarial and actual utilization perspectives are of interest, they are both reported.

In analyzing the economic merit of a second-opinion consultation program, one should focus on total group (i.e., cohort) utilization and cost patterns. Within any group, savings can emerge both from changes in medical care utilization patterns among those who utilized care and from the avoidance of care. The latter phenomenon can only be seen if total group (i.e., cohort) patterns are analyzed. The same rationale dictates that group-wide patterns of work loss and restricted-activity days are pertinent for the current study. Consequently, the cost–benefit analysis focuses primarily on cohort-specific data.

Study Population

The study sample was drawn from the population of all covered individuals who were advised to have elective surgery during the period from January 1, 1977 to December 31, 1978 (n = 2,284). Of this group, 366 individuals did not have their need for surgery confirmed by a consultant surgeon. A comparable number of individuals whose need for surgery was confirmed by a consultant surgeon were randomly selected from the total number of confirmed individuals (n = 1,918).** Data could not be obtained for 24 not-confirmed individuals; thus, the final study population consisted of 342 not-confirmed and 342 confirmed individuals.††

The median age category for both the confirmed and not-confirmed groups is 45–64 years. Approximately 7 per cent of the total study population were in age category 15–24 years, 37 per cent were in age category 25–44 years, 53 per cent were in age category 45–64 years and 3 per cent were 65 years of age or older. Fifty-two per cent of the study population were male, and 48 per cent were female. Seventy-five per cent of the individuals included in the study were married, 13 per cent were single and 12 per cent were divorced or separated. Approximately 67 per cent were employed within the six month period following the second-opinion consultation, and an additional 25 per cent were homemakers (housewives). Only 1 per cent of the study population were unemployed. The remaining study participants (6 per cent) were students, retirees or permanently disabled. Approximately half of the individuals in both the confirmed and not-confirmed populations reported having completed less than 12 years of formal schooling, and an additional 14 per cent indicated that they did not know the extent of their educational attainment. Only 2 per cent had completed college. Two thirds of the employed respondents in both the confirmed and not-confirmed groups reported weekly earnings in the $200–299 range.

Ten surgical procedures account for approximately 60 per cent of the initial recommended surgeries: eye surgery except cataract (9.3 per cent); mass, cyst or lesion excision other than varicose veins (8.8 per cent); hernia repair (8.8 per cent); hysterectomy (7.4 per cent); dilation and curettage (5.4 per cent); incision, excision and suture for mass, cyst or lesion of the skin (4. 8 per cent); repair of deviated septum (3.9 per cent); foot surgery except bunionectomy (3.8 per cent); removal of cataract (3.6 per cent); and mastotomy, mastectomy and other operations of the breast (3.5 per cent).

Each person included in the cost–benefit analysis was followed for a period of one year following their second-opinion

---

** The data pool available for analysis did not permit a case-by-case match by diagnosis.

†† This sample size is greater than that required at $p \leq 0.05$ to note signifiant program savings.

consultation. As entry into the study spanned a two-year period (January 1, 1977 to December 31, 1978), cost data spanned a three-year period—1977 through 1979. All 1978 and 1979 cost data were deflated to a 1977 base, utilizing a 10 per cent discount rate.

Data Base

Data Sources and Proxies

Data utilized in this study were drawn primarily from welfare fund medical claim records and secondarily from telephone-administered questionnaries administered to all study participants one year after the anniversary date of the second-opinion consultation. Interviews were utilized to ascertain the amount of work loss and restricted-activity days, and out-of-pocket medical care costs. Data on medical care utilization were also gathered through the questionnaire and used in the analysis in instances where the person's claim file was incomplete. Data on the cost of the program were obtained from records maintained by the fund.

In generating the data base for this study, extreme care was exercised to ensure that medical utilization, work-loss and restricted-activity data related *only* to the condition for which surgery was originally recommended were included. Medical utilization, employment and restricted-activity data resulting from other causes were *excluded* from the study.

Ninety-two per cent of the study population (629 people) responded to the cost–benefit questionnaire. As medical claim form data were available for nonrespondents, it was decided to include them in the study and to proxy the missing data. Where precise information was not reported for the cost of office-based physician visits, the price of such visits was proxied at $30/visit. This proxy value was selected as it represented the prevailing office visit charge in the New York City area during the time of the study. In all cases, hospital costs were estimated based on the prevailing Blue Cross and Blue Shield of Greater New York per diem reimbursement. Hospital charge data were not used since charges overestimate the true cost of rendering hospital care. As the Blue Cross policy offered by the welfare fund provides full coverage for the first 21 days of a hospital stay and half coverage for the succeeding 180 days, the Blue Cross payment for care beyond the 21st day was doubled. Charges for medically related services not covered by Blue Cross, such as the cost of blood, were added to the reimbursement revenue to obtain the aggregate cost of hospital care.

In cases in which a specific drug cost was not provided, but in which the respondent reported the type of drug prescribed and the frequency and duration of intake, drug costs were proxied using prices listed in the *Lilly Drug Directory*. In the case of eye surgery, the cost of eye drops was proxied at $1 per month, the cost of eye glasses at $75/pair and the cost of contact lenses at $250/pair. These estimates were based on expert medical opinion. If direct proxies could not be obtained, mean values for the particular data element based on information provided by the applicable population group (i.e., confirmed and not-confirmed) was used as an indirect proxy.

Data on individuals' earnings were required to calculate the value of incurred work losses. Where necessary, wage data for Building Service union members were proxied based on collective bargaining hourly wages for the pertinent occupational category. For non-Building Service spouses, proxies were derived, where necessary, from Bureau of Labor Statistics regional wage reports.

The price used to value homemaker (housewife) time was based on the market valuation of homemaker services generated by Wendyce Brody.[9] Brody estimated the annual value of housewife services in

1972 as $4,705. This estimate was based on market wages prevailing in the Syracuse, New York area. Bureau of Labor Statistics wage level indices were used to: a) adjust the data to reflect wage patterns prevailing in the New York City region and b) inflate 1972 data to reflect market values prevailing during the current study period.

The prevailing minimum wage was used to value the restricted-activity time of students, retirees, permanently disabled, and laid-off or unemployed individuals. If an individual's employment status changed during the study period (i.e., person retired, was laid off, etc.), the value of his/her lost time (i.e., actual or minimum wage) was determined by the employment status at the time the loss occurred.

Biases Inherent in the Data Base

Factors unique to the welfare fund study population and to the New York City health care system introduce a downward bias in some of the cost data utilized in this study. Specifically, surgeon and physician visit costs may be biased in a downward direction due to three factors: 1) the welfare fund's surgical fee negotiation program, 2) the availability of "panel" surgeons, and 3) the billing practices of the welfare fund's health center. The surgical fee negotiation program attempts to obtain reductions in surgical fees in cases where the surgeon accepts assignment and his fee exceeds the welfare fund's stipulated reimburement for any procedure.[10] Under this program, surgeons are initially requested to accept the welfare fund's reimbursement as payment in full or, if this is not acceptable, to reduce their fee voluntarily. In 7.7 per cent of the surgical cases included in the cost–benefit study, the surgeon's fee initially exceeded the welfare fund's allowable scale, and negotiation occurred.

Some individuals electing surgery may have used a "panel" surgeon. If a beneficiary, having gone for a second-opinion consultation, does not have a surgeon to perform the elective surgery, the welfare fund will arrange for a "panel" surgeon to perform the surgery. Panel surgeons are board certified and agree to accept the welfare fund's designated reimbursement. Furthermore, panel members are not permitted to bill the welfare fund or the beneficiary for routine postoperative medical care (5.7 per cent of the people electing surgery had their surgery performed by a panel surgeon).

Building Service Local 32B-J also operates a health center for its members. Fees at the health center are lower than those prevailing in the city at large (e.g., the cost of a physician visit is only $25). Furthermore, if a patient receives medical care there, the welfare fund will pay only 80 per cent of the medical bill after the annual $100 deductible is met. Health center physicians quite often fail to bill the patient for the remaining 20 per cent.

New York City has the largest network of municipal hospitals in the nation. Although municipal hospitals are frequently thought of as places that treat only the poor, they do render a substantial amount of care to the working population of the city. Two biases are introduced if care was obtained in the municipal system. Cost of outpatient department care may be biased downward as charges for such care are based on the patient's ability to pay. Furthermore, when a patient is hospitalized in a municipal hospital, he/she is treated by a salaried surgeon who is not permitted to bill the patient for this care. The cost of these services is presumably part of the total hospital bill. The same applies to the cost of anesthesiology.

A comparable situation exists when a patient is hospitalized in one of the city's major teaching hospitals. If the patient is placed on the teaching, as opposed to the private, service, a separate surgeon's charge is not forthcoming because the cost of medical and surgical care is presumably included in the aggregate hospital bill. Finally, in a number of cases, anesthesiologists did not bill the patient directly but included their fees as part of the surgeon's bill. In such cases (n = 61), the cost of surgical care is overestimated and the cost of anesthesiology care is underestimated.

The net effect of all these biases probably leads to an *underestimate* of the true savings of the second-opinion consultation program. The major impact of these biases is to reduce the cost of surgery, thereby underestimating program benefit. One should recognize that comparable biases would result in any study focused either on an urban lower-class working population or on special population groups such as Medicare and Medicaid patients, whose payment for care may be below prevailing physician charges.

Program Savings

Nine medical resource cost categories were delineated for this study: physician visits; hospital utilization; surgical fees for the original surgery; anesthesia fees for the original surgery; surgical fees for related surgery‡‡; anesthesia fees for a related surgery; diagnostics and laboratory; therapies and nursing; and drugs and prostheses. In addition, data are available on the number of physician visits, number of hospitalizations, total number of days of hospital care rendered and number of surgeries performed.

Medical resource utilization patterns displayed by the not-confirmed and confirmed groups are reported in Table 1. On

TABLE 1. Medical Resource Utilization Profiles

	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Utilizers		Cohort			Utilizers	
Category	Total	Mean	Standard Deviation	Number	Mean	Total	Mean	Standard Deviation	Number	Mean
Number of physician visits	1,206	3.5	7.1	185	6.5	1,077	3.1	6.6	175	6.0
Number of hospital stays	325	0.9	0.6	298	1.1	151	0.4	0.6	133	1.1
Length of stay—all hospitalizations	2,130	6.2	7.7	298	7.2	1,206	3.5†	7.0	133	9.0
Number of surgeries*	396	1.2	0.8	304	1.3	193	0.6†	0.9	132	1.5

* This category includes the surgical procedures for which the individual received a second-opinion consultation (called original surgery) and all other surgical episodes related to the medical condition for which a second-opinion consultation was sought.

† Statistically significant from the mean value reported for the confirmed group at $p \leq 0.01$.

---

‡‡ In most cases, related surgery is a surgical procedure associated with the diagnosis for which a second opinion was sought. It is not, however, the procedure that was initially recommended. For example, the procedure initially recommended was dilation and curettage. This procedure was performed, and subsequently a hysterectomy was required. In other cases, related surgery refers to a diagnostic procedure, such as scoping, that is performed in a hospital under anesthesia and is done as a prelude to surgery, e.g., cystoscopy prior to a transurethral resection (prostatectomy).

both a cohort and utilizer basis, fairly comparable patterns emerge with regard to number of physician visits. Individuals reporting the use of physician care had approximately six visits in the follow-up year. On a cohort basis, utilization falls to approximately three visits per person per year. The number of hospital stays for those individuals reporting hospitalization was the same—1.1 stays per person. While hospital utilization among the not-confirmed group appears to be much lower on a cohort basis (0.4 versus 0.9 hospital stays), this difference fails to attain statistical significance at $p \leq 0.05$. Statistically significant differences do emerge with regard to cohort profiles for length of hospital stay and number of surgeries. For both of these phenomena, average use by the entire confirmed group was almost double that of the entire not-confirmed group—6.2 versus 3.5 hospital days, and 1.2 versus 0.6 surgeries.

Cost profiles generated from data reported by confirmed and not-confirmed individuals indicate that, on average, individuals in the confirmed group spent more money on medical care than their counterparts in the not-confirmed group, as can be seen from the data displayed in Table 2. Average cost per utilizer in the confirmed group was $2,429, as compared with $1,850 for the not-confirmed group. For those individuals utilizing medical care, mean cost of physician care, anesthesiology (related surgery), diagnostics and laboratory, and therapies and nursing was higher for the confirmed than for the not-confirmed group. For the remaining five categories—hospital care, surgical fees (original surgery), anesthesiology fees (original surgery), surgical fees (related surgery), and drugs and prostheses—the reverse was true.

Cohort-wide cost patterns also indicate that average cost per person in the not-confirmed group was less than that reported for the confirmed group. Total medical care expenditures incurred by the entire not-confirmed group equaled $425,538, contrasted with $787,294 for the entire confirmed group. Viewed from another perspective, average medical care cost per cohort member was $1,244 for the not-confirmed group and $2,302 for the confirmed group. As expected, two categories account for the majority of medical care expenditures for both cohorts. These categories are hospital utilization and surgical fees (original surgery). Jointly they account for 84 per cent of the expenditures incurred by the confirmed group and 79 per cent of the expenditures incurred by the not-confirmed group. On a per-cohort-member basis, differences in mean expenditures for hospital utilization were $635 ($1,348 − $713), and differences in mean expenditures for surgical fees for the original surgery were $324 ($601 − $277). It is interesting to note that statistically significant differences in total group expenditures appear for two additional medical resource categories—anesthesiology fees (original surgery) and therapies and nursing. In both cases, expenditures incurred by the confirmed group exceeded those incurred by the not-confirmed group.

Seven productivity-loss categories were delineated for this study: work/home/school loss due to days spent in the hospital (labeled "hospitalization"); posthospitalization regular work (labeled "main job") loss; overtime work loss§§; work loss at any additional jobs; posthospitalization restricted-activity days incurred by students, retired, disabled or unemployed individuals; and work-loss days incurred by relatives (other than spouses) of patients who had to attend to the needs of the patient. In addition, data were gathered on expenditures for outside household help that had to be hired to substitute for an ill housewife, and time loss (i.e., travel, wait-

---

§§ If a spouse incurred work-loss days to attend to the needs of the patient, this loss was included in the main job and overtime—main job categories.

TABLE 2. Medical Resources Cost Profiles

	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Utilizers		Cohort			Utilizers	
Category	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)
Physician visits	31,074	91	207	185	166	27,902	82	185	175	156
Hospitalization	460,979	1,348	2,013	298	1,545	243,819	713*	1,604	133	1,821
Surgical fees—original surgery	200,068	601†	496	296	676	92,844	277*‡	479	120	774
Anesthesiology fees—original surgery	50,342	185§	131	217	214	20,692	68*‖	134	86	230
Surgical fees—related surgery	15,661	46	217	24	652	21,090	62	329	23	901
Anesthesiology fees—related surgery	4,343	13	61	19	229	4,827	14	62	20	221
Diagnostics and laboratory	10,446	31	166	49	212	5,361	16	54	50	106
Therapies and nursing	7,447	22	162	13	570	1,368	4#	38	6	223
Drugs and prostheses	6,934	20	92	78	85	7,635	22	173	65	115
Total cost¶	787,294	2,302	2,741	322	2,429	425,538	1,244*	2,427	228	1,850

* Statistically significant from the mean value reported for the confirmed group at $p \le 0.01$.

† The number of observations used to calculate this mean was 333. Nine individuals in this group did not have separate, identifiable surgical bills.

‡ The number of observations used to calculate this mean was 334. Eight individuals in this group did not have separate, identifiable surgical bills.

§ The number of observations used to calculate this mean was 272. Seventy individuals in this group either did not incur any anesthesiology costs or did not have separate bills for anesthesiology.

‖ The number of observations used to calculate this mean was 306. Thirty-six individuals in this group either did not incur any anesthesiology costs or did not have separate bills for anesthesiology.

# Statistically significant from the mean value reported for the confirmed group at $p \le 0.05$.

¶ The number of utilizers in the total cost row exceeds the number in any previous row. This is due to the fact that any individual may report data for more than one category, but not for all categories. Thus the statistic listed in the total cost row indicates the number of individuals who reported *any* expense.

TABLE 3. Productivity Losses (in days)

Category	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Reported		Cohort			Reported	
	Total	Mean	Standard Deviation	Number	Mean	Total	Mean	Standard Deviation	Number	Mean
Hospitalization	2,130	6.2	7.7	296	7.2	1,209	3.5*	7.0	131	9.2
Main job	4,940	14.5	21.8	158	28.3	2,115	6.2*	20.1	85	23.4
Overtime (in hours)	556	1.6	8.4	19	27.1	401	1.2	7.7	13	25.4
Additional job	28	0.1	1.2	2	13.0	0	—	--	0	—
Homemaker	1,846	5.4	21.9	46	39.4	902	2.6	17.0	24	35.9
Restricted activity	1,097	3.2	22.3	18	55.9	1,267	3.7	26.9	12	99.8
Relative work loss	400	1.2	12.3	24	15.5	83	0.2	1.4	11	6.6

* Statistically significant from the mean value of the confirmed group at $p \leq 0.01$.

ing and examination time) and travel costs incurred for obtaining medical care in the year following the second-opinion consultation.

Productivity-loss patterns are presented in Table 3. On average, individuals in the confirmed group who reported actual productivity losses sustained 7.2 work/home/school loss days due to hospitalization and 28.3 main job work-loss days in the posthospitalization period. Not-confirmed individuals sustained 9.2 work/home/school loss days due to hospitalization and 23.4 posthospitalization main job work-loss days. Housewives (homemakers) in the confirmed group reported an average home loss of 39.4 days, while those in the not-confirmed group reported home losses averaging 35.9 days.

When viewed from the perspective of the entire cohort, much greater mean productivity losses emerge. The entire not-confirmed group averaged approximately half as many work/home/school loss days due to hospitalization and posthospitalization work-loss days as the entire confirmed group. Averaged over the entire cohort, confirmed group members lost 6.2 work/home/school days due to hospitalization, while not-confirmed group members lost only 3.5 days. Furthermore, confirmed individuals averaged 14.5 posthospitalization work-loss days, while not-confirmed individuals lost only 6.2 work days. Both the hospitalization and posthospitalization work-loss patterns are statistically significant at $p \leq 0.01$. Differences with regard to the remaining productivity-loss categories (i.e., overtime loss, additional job loss, homemaker loss days, restricted-activity days and relative work-loss days) were not statistically significant at $p \leq 0.05$.

The value of the total productivity loss sustained by the confirmed and not-confirmed groups is reported in Table 4. The entire not-confirmed group productivity loss was $175,621, contrasted with $348,656 for the entire confirmed group, resulting in a statistically significant program saving of $173,035. The three major productivity cost components, accounting for over 80 per cent of the total cost incurred by each group, were posthospitalization main job work loss, hospitalization work/home/school loss, and posthospitalization home-loss days incurred by housewives. In each of these cases, average group losses ($539 versus $243, $202 versus $105 and $144 versus $69) were significantly different at $p \leq 0.05$ or lower.

By combining medical resource expenditures and productivity-loss patterns, one obtains the total economic costs incurred by the not-confirmed and the confirmed groups. Total economic costs incurred by the not-confirmed cohort in the one-year follow-up period were $601,159 ($425,538 in medical care costs and $175,621 in productivity losses), as compared with $1,135,950 ($787,694 for medical care and $348,656 in productivity losses) incurred by the confirmed group. Subtracting the costs incurred by the not-confirmed cohort from those incurred by the confirmed cohort yields a total program saving (i.e., benefit) of $534,791. Sixty-eight per cent of this total saving resulted from decreased use of medical resources and 32 per cent from decreased productivity losses.

Program Cost

Program costs consist of the administrative costs incurred to operate the program, the expenditures for the second-opinion consultation visits and any ancillary tests ordered as part of the consultation, and out-of-pocket (i.e., travel) and opportunity (i.e., wage loss) costs incurred in the process of obtaining the consultation.

The costs of administering the second-opinion program during 1977 were estimated from records maintained by the Building Service Local 32B-J welfare fund. Personnel costs totaled $29,350, of which $19,350 were salary and fringe benefit

TABLE 4. Cost of Productivity Loss

	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Reported		Cohort			Reported	
Category	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)
Hospitalization	69,030	202	249	293	236	36,063	105*	232	129	278
Main job	184,193	539	814	158	1,056	82,953	243*	934	85	899
Overtime	4,575	13	76	19	224	2,846	8	56	12	221
Additional job	1,121	3	55	2	530	0	—	—	0	—
Homemaker	49,173	144	581	46	1,049	23,751	69†	453	24	954
Restricted activity	16,566	48	361	18	906	20,023	59	435	12	1,616
Relative work loss	13,333	39	283	20	437	2,352	7†	44	11	185
Outside household help	882	3	38	2	437	735	2	28	2	367
Time lost for obtaining medical care	6,735	20	25	252	22	4,681	14*	23	171	22
Cost of travel for obtaining medical care	3,048	9	17	241	11	2,217	6	16	159	11
Total cost	348,656	1,019	1,385	315	1,030	175,621	514*	1,170	207	802

* Statistically significant from the mean value of the confirmed group at $p \leq 0.01$.

† Statistically significant from the mean value of the confirmed group at $p \leq 0.05$.

costs for 1.5 (FTE) intake officers, and $10,000 were allocated as the stipend of the welfare fund's medical advisor. Nonsalary costs totaled $10,000, of which $7,200 are imputed rental costs, and $2,800 represent actual incurred costs for office expenses. A separate study was not conducted to ascertain 1978 costs, which would have had to be deflated to a constant (1977) dollar level. As no change occurred in the administrative structure of the program during 1978, 1977 costs were used as a proxy for 1978 deflated costs. As program intake spanned a two-year period (January 1, 1977 through December 31, 1978), it is estimated that total administrative costs were $78,700.

During the two-year study period, 2,284 individuals obtained second-opinion consultations and 24 individuals obtained an additional (i.e., third) opinion. On average, the costs of the second-opinion consultation and associated ancillary procedures was $43.95, and the average cost of third-opinion consultations and associated ancillary procedures was $47.42. [Both of these cost statistics are in constant (1977) dollars.] Based on the number of consultations rendered, total expenditures for all consultations were $101,520.

Detailed data were not gathered on travel costs and opportunity costs incurred in obtaining second-opinion consultations. It is assumed that travel to and from the consultant's office as well as the examination took, on average, two hours. The average hourly wage prevailing for the study population, $5, was used to derive an aggregate travel and opportunity cost estimate of $23,080.

Combining the three program-cost elements yields a total program-cost estimate of $203,300. Fifty per cent of the costs were incurred for medical expenses associated with the second-opinion consultation, 39 per cent of the costs were accounted for by administrative expenses, and 11 per cent of the costs reflect out-of-pocket and wage loss incurred to obtain the consultation.

Cost–Benefit Calculations

Based on the experience of the not-confirmed and confirmed groups, which include both individuals who complied with the recommendation of the second-opinion consultant and those who did not, one obtains a benefit–cost ratio of 2.63. The components of this ratio are displayed in Table 5. This benefit–cost ratio implies that for every $1.00 of program cost incurred, $2.63 of benefits accrue.[11] An alternate evaluation statistic is the program's net present value (NPV). This is obtained by subtracting program cost from program benefit. The NPV for the second-opinion consultation program is $331,491. This indicates that net benefit (i.e., total benefit minus total cost) equaled $331,491 for the program under study. Dividing net benefit by the total number of program participants (2,284) yields a net benefit per participant of $145.14. Thus, on average, for every individual who is told to obtain a second-opinion consultation (a population pool much larger than the number not-confirmed), one can expect to save $145.14.

The perspective adopted in the current study recognizes benefits from both reduced utilization of medical resources and reductions in productivity loss. If only medical resource utilization that can lead to a direct reduction in medical care expenditures and cost borne directly by the program had been considered in the analysis, the ensuing benefit–cost ratio would have been 2.00. Even from this (narrow) perspective, the second-opinion consultation program must be considered

---

[11] In order to test the sensitivity of this finding to the selected discount rate, 10 per cent, the entire analysis was repeated using alternately rates of 15 per cent and 6 per cent. The resulting benefit–cost ratios were 2.53 and 2.72, respectively.

TABLE 5. Cost–Benefit Statistics

Component	Value
Total costs incurred by the confirmed group (B_e)	$1,135,950
Total costs incurred by the not-confirmed group (C_e)	$ 601,159
Program cost (U_e)	$ 203,300
Number of participants (P)	2,284

Cost–Benefit Equations

$$\text{Benefit/Cost} = \frac{B_e - C_e}{U_e} = \frac{1{,}135{,}950 - 601{,}159}{203{,}300} = \frac{534{,}791}{203{,}300} = 2.63$$

$$\text{Net Present Value} = (B_e - C_e) - U_e = 534{,}791 - 203{,}300 = \$331{,}491$$

$$\text{Net Present Value/Participant} = \frac{(B_e - C_e) - U_e}{P} = \frac{331{,}491}{2{,}284} = \$145.14$$

cost-beneficial in that for every $1.00 of cost incurred, $2.00 of benefit (reduced medical resource utilization) occurred.[##]

Summary and Discussion

An economic evaluation of the mandatory second-opinion consultation program administered on behalf of the Building Service Local 32B-J welfare fund indicates that substantial net savings were reaped both by the welfare fund and its covered beneficiaries. Comparing the medical utilization histories of the not-confirmed and the confirmed groups during the first year after receipt of the consultation indicates that the not-confirmed group spent $361,756 less on medical care than the confirmed group. Furthermore, due to the avoidance of surgery, the not-confirmed group exhibited fewer work-loss, home-loss and restricted-activity days. This saving, referred to as a productivity gain, was valued at $173,035. Jointly, savings from reduced medical care utilization and increased productivity equalled $534,791.

Total costs incurred by the second-opinion consultation program during the two-year study period (1977–1978) were $203,300. Total expenditures for consultations and associated ancillary procedures

---

[##] The computation of medical resource savings used throughout this study is based on a long-run perspective in which all costs are considered to be variable. In the short-run, some costs are fixed and will be incurred even when utilization declines. This is particularly true in the case of hospital care. As agreement does not exist as to the percentages of total costs that are fixed and variable, respectively, we assume that 50 per cent of costs are fixed and 50 per cent variable. Utilizing this assumption, short-run savings due to decreased hospital use decline from $217,610 (i.e., $460,979 – $243,819; line 2, Table 2) to $108,580, thereby reducing the aggregate benefit–cost ratio from 2.63 to 2.10, and the benefit–cost ratio based exclusively on reduced medical resource utilization declines from 2.00 to 1.25. It is our opinion, however, that a long-run perspective, in which all costs are variable, is the proper one for policy analysis.

equalled $101,520. Administrative costs were $78,700, and estimated out-of-pocket and opportunity costs were $23,080.

Merging the estimates of program benefit and cost yields a benefit–cost ratio of 2.63. That is, for every $1.00 of cost incurred, $2.63 of benefits occurred. Based exclusively on this "pay-back ratio," one must conclude that there are definite economic benefits to be derived from a second-opinion consultation program. While this evaluation did not focus on noneconomic benefits, the majority of individuals responding to questionnaires volunteered the information that they were feeling better during the postconsultation period.

Other evaluations of second-opinion consultation programs have suggested that the receipt of a second-opinion consultation is cost beneficial.[5,6,11] The results of the current study add further weight to this contention. Furthermore, the results of the current study should be viewed as more definitive, as the methodology and data base used in this study overcome many of the shortcomings inherent in previous evaluations of the economic potential of second-opinion consultation programs.

The first published evaluation of a second-opinion consultation program focused on the intial 20 months of Blue Cross and Blue Shield of Greater New York's voluntary second-surgical-opinion program. One of the major findings reported in this evaluation, based on the experience of over 1,500 program participants, was an estimated gross savings of $600,000 in health care costs. After deducting the costs of administering the program, which included relatively high start-up costs, the evaluators estimated a net savings due to the program of a quarter of a million dollars.[11] It should be recognized that this estimate excludes any productivity gains resulting from the avoidance of work-loss or home-loss days and is based exclusively on assumed gross savings from the avoidance of surgery. Thus, from a societal perspective, the reported savings are underestimates. In generating the assumed savings resulting from the avoidance of surgery, the relative experience of a "control" group similar to that specified in this study was not utilized. Therefore, savings reported from the avoidance of surgery must be viewed as gross savings that were not offset by expenditures incurred by individuals not electing surgery, and as such, these reported savings are overestimates of true medical resource savings.

A recently published evaluation of Massachusetts' mandatory consultation program for elective surgery for Medicaid recipients estimated that net statewide savings to Medicaid during the first year of program operation were $856,500, and when benefits are compared with costs, the program exhibits a benefit–cost ratio of 3.9.[5] It should be noted that this study considered program costs and savings strictly from Medicaid's perspective. If costs were borne by other parties, or if benefits accrued to other parties, they were not included in the analysis. Furthermore, data on medical care expenditures were based on estimated Medicaid payments rather than claim files data, which were used in the current study. Productivity gains or losses were omitted from the analysis of program benefit, and gross estimated savings from foregone surgery were offset only by the cost of other surgery, if other surgery was recommended by the consultant. Nonsurgical expenses, such as those for physician visits and drug therapy, were not estimated for individuals foregoing surgery and were not subtracted from estimated gross savings. As in the case of the Blue Cross and Blue Shield of Greater New York study, the reported program savings and cost–benefit ratio were biased by these omissions.

Gertman et al. present an alternate estimate of the economic value of the Massachusetts program, based exclusively on the experience of 49 patients living in the metropolitan Boston area who received second opinions reversing the recommendation for a hysterectomy. In this study, program benefit was also estimated solely by projecting savings from changes in the amount of surgery performed. Based on this procedure, a benefit–cost ratio of 2.27 was reported. Applying this methodology to the other seven procedures encompassed by the Massachusetts program,¶¶ Gertman et al. speculate that the benefit–cost ratio would be only 1.11.[6] the shortcomings inherent in the first study of the Massachusetts Medicaid program apply with equal force to the Gertman et al. study.

The economic evaluation presented in the current study does not suffer from the shortcomings inherent in these sudies. Our perspective recognizes all social costs, not just costs germane to one party such as a Medicaid program. The definition of program benefit used in the current study encompasses both medical resource and productivity savings. Furthermore, the medical resource savings reported in this study are based on actual data maintained in a welfare fund's medical claims files. Cost estimates were used only when proxy values were needed.

While the selection of a "control" group for the current study represents a distinct methodological improvement over previous evaluation efforts in this subject area, it is recognized that an optimal control group was not used. Data availability limitations and the widespread knowledge of the merit of "second opinions" precluded the selection of a different control group. Thus, a key question raised by critics of a second-opinion consultation program, namely, "Does the program recruit individuals to surgery?" cannot be answered by our analysis. An exhaustive search of the literature failed to turn up any data on the proportion of individuals who decide not to undergo elective surgery without the benefit of additional information provided by a second-opinion consultation. Discussions with various welfare fund administrators in the New York City area that are either self-insured, self-administer their health benefits program or have traditional third-party carrier relationships indicate that the surgical utilization experience of their membership and covered dependents did *not* increase after the adoption of a second-opinion consultation program. In light of this development we proceeded with the use of the control group selected for this study.

All economic evaluations of second-opinion consultation programs have been based on a limited follow-up period. Evidence gathered in the course of this and other evaluations indicates that approximately 90 per cent of all individuals who will elect surgery after a second-opinion consultation will do so within six months of the consultation. Thus, the vast majority of medical care costs and productivity losses will occur within a year following the receipt of the consultation. Furthermore, when program participants have been followed for longer periods (i.e., two years), no evidence has been uncovered to indicate that any new medical complications occur.

The cost–benefit analysis presented here indicates that the second opinion consultation program administered on behalf of the Building Service Local 32B-J welfare fund returned $2.63 in benefits for every $1.00 of cost incurred. Had the program required that adherence to the consultant's recommendation be a prerequisite for full medical coverage, as is the case with the second-opinion program offered by the Prudential Insurance Company of America; had it covered a labor force commanding higher wage levels; and had it been targeted to selective diagnoses, the benefit–cost ratio would have exceeded

---

¶¶ These procedures are tonsillectomy and adenoidectomy, cholecystectomy, hemorrhoidectomy, laminectomy or spinal fusion, meniscectomy, submucous resection or nasal septal repair, and excision of varicose veins.

2.63. Support for the first contention comes from one of the substudies (not reported here) conducted as part of our overall research effort. When the analysis was limited solely to individuals who followed the recommendation of the consultant (cohorts 1 and 7 in the Appendix table), the reported benefit–cost ratio was 3.18, indicating a saving of $3.18 for every $1.00 expended. As noted in the "Study Population" section of this article, the median weekly salary for Building Service employees was in the $200–$299 range, implying that median hourly earnings are in the $5–$8 range. Had the covered employees commanded hourly wages in the $12–$16 range, as in the case for carpenters and teamsters, productivity savings would have been higher. This is the case because work-loss days avoided would have been valued at the higher wage level. Greater productivity savings would automatically lead to a higher benefit–cost ratio. Finally, there is reason to believe that economic benefits could be maximized by confining second-opinion consultation programs to those diagnoses that have the highest nonconfirmation rates. While additional research is needed to identify these diagnoses, the data reported in the evaluation of the Massachusetts program indicate substantial differences in net savings for each of the eight procedures relevant to the Massachusetts Medicaid program.[5]***

Conclusion

Second-opinion consultation programs represent the first concerted cost-containment effort targeted directly at consumers and individual providers of medical care. Prior cost-containment efforts have focused almost exclusively on the institutional providers of care. While early efforts at controlling cost via rate regulation[12–14] and utilization review[15–19] were not overly successful, these efforts are becoming more successful as the respective programs have matured.[20,21] To date, a comprehensive evaluation of rate regulation from a cost–benefit perspective has not been undertaken. While it has been shown that rate regulation reduces the rate at which hospital costs increase in a number of large industrial states, one is not sure whether the gains are being offset by the cost of the rate setting activity. Furthermore, it should be evident that rate regulation will succeed only in those states that have the political will to enforce and sustain such a system.

It is reasonable to assert that the control of health care costs requires a multiplicity of efforts. Although control of institutional costs is beginning to yield positive results, one must also develop programs that are directed toward the individual consumers and their physicians, who play a dominant role in determining the volume and type of medical care services that are consumed. To date, the PSRO program, which by design entails a *retrospective* review, has offered few direct challenges to the individual physician and does not attempt to interact directly with consumers. Second-opinion consultation programs are currently the only major *prospective* effort to influence the consumer and, through the consumer, to influence physician behavior indirectly. As such programs can both reduce surplus surgery and effectively contain costs, they merit inclusion in an array of public or private sector cost-containment efforts.

References

1. McCarthy EG, Widmer GW. Effects of screening by consultants on recommended elective surgical procedures. N Engl J Med 1974;291:1331.

2. Grafe WR, et al. The elective surgery second opinion program. Ann Surg 1978;188:323.

*** Reporting on savings estimated for patients whose need for surgery was reviewed by the Bay State Health Care Foundation, Martin et al. estimate an aggregate benefit–cost ratio of 4.2. When procedure-specific estimates were made, the benefit–cost ratio ranged from a low 0.9 for cholecystectomy to a high of 7.8 for excision of varicose veins.

3. McCarthy EG, Finkel ML. Second consultant opinion for elective surgery programs: outcome status over time. Med Care 1978;16:984.

4. McCarthy EG, Finkel ML. Second consultant opinion for elective gynecologic surgery. Obstet Gynecol 1980;56:403.

5. Martin SG, et al. The effect of a mandatory second opinion program on Medicaid surgery rates: an analysis of the Massachusetts consultation program for elective surgery. Boston: Department of Public Welfare, Commonwealth of Massachusetts, 1980. (Mimeographed).

6. Gertman PM, et al. Second opinions for elective surgery: the mandatory medicaid program in Massachusetts. N Engl J Med 1980;302:1169.

7. Joffe J, Schachter M. Program for elective surgical second opinion: surgical experience of program participants, 1976-77. (PRESSO Report No. 3). New York: Health Affairs Research Department, Blue Cross and Blue Shield of Greater New York, 1980. (Mimeographed).

8. McCarthy EG, Finkel ML, Ruchlin HS. Second opinion elective surgery. Boston: Auburn House, 1981.

9. Brody WH. Economic value of a housewife. Research and Statistics Note 9-1975. DHEW Publication No. (SSA) 75-11701, August 28, 1975.

10. Kamons AS, et al. Cost containment by a third party payer: negotiations of surgical fees. Employee Benefits Journal 1980;5:2.

11. Joffe J. Programs for elective surgical second opinion evaluation: preliminary findings, January 1976–August 1977. (PRESSO Report No. 1). New York: Health Affairs Research Department, Blue Cross and Blue Shield of Greater New York, 1978. (Mimeographed).

12. Berry RE. Prospective rate reimbursement and cost containment: formula reimbursement in New York. Inquiry 1976;13:288.

13. Hellinger FJ. Prospective reimbursement through budget review: New Jersey, Rhode Island, and Western Pennsylvania. Inquiry 1976;13:309.

14. Baker FD. Prospective rate setting in Washington State. Topics in Health Care Financing 1976;3:57.

15. Flashner BA, et al. Professional standards review organizations: analyses of their development and implementation based on a preliminary review of the hospital admission and surveillance program in Illinois. JAMA 1973;223:1473.

16. Brian E. Foundation for medical care control of hospital utilization: CHAP—A PSRO protype. N Engl J Med 1973;288:878.

17. Westphal M, Frazier E, Miller MC. Changes in average length of stay and average charges generated following institution of PSRO review. Health Serv Res 1979;14:253.

18. Lave JR, Leinhardt S. An evaluation of a hospital stay regulatory mechanism. Am J Public Health 1976;66:959.

19. Brook RH, Williams KN. Evaluation of the New Mexico peer review system, 1971 to 1973. Med Care 1976;14(Supplement):

20. Biles B, Schramm CJ, Atkinson JG. Hospital cost inflation under state rate-setting programs. N Engl J Med 1980;303:664.

21. Health Care Financing Administration, Office of Research, Demonstrations, and Statistics. Professional Standards Review Organization, 1979 Program Evaluation. HCFA Publication No. 03041. Washington, D.C.: U.S. Government Printing Office, 1980.

Call for Abstracts
American Public Health Association
1982 Annual Meeting

The Medical Care Section of the American Public Health Association is calling for abstracts of research dealing with the organization, financing, and utilization of health care. These abstracts must describe projects that will be completed by June 1982. The papers selected will be presented at the 110th Annual Meeting to be held November 15–18, 1982, in Montreal, Canada.

Abstracts must use the standard form appearing in the January and February 1982 issue of *The Nation's Health* or available from the contract below. All submissions to the Medical Care Section must be postmarked no later than March 12, 1982. Please submit to: Nelda McCall, Director, Health Policy Research, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025.

Appendix
Select Medical Resource Utilization and Cost Profiles

Cohort	Resource Category	Cohort Total	Cohort Mean	Cohort Standard Deviation	Utilizers Number	Utilizers Mean
1	Confirmed, surgery elected (N = 305)					
	Physician visits	1,131	3.70	7.3	171	6.6
	Hospital stays	324	1.10	0.5	297	1.1
	Number of surgeries	395	1.30	0.7	305	1.3
	Total cost	$782,612	$2,566	2,788	305	$2,550
2	Confirmed, surgery not elected (N = 37)					
	Physician visits	75	2.00	4.6	17	4.3
	Hospital stays	1	0.03	0.2	1	1.0
	Number of surgeries*	1	0.03	0.2	1	1.0
	Total cost	$4,682	$127	342	17	$263
3	Not-confirmed, no surgery (N = 213)					
	Physician visits	552	2.60	7.0	92	5.9
	Hospital stays	12	0.60	0.2	11	1.0
	Number of surgeries*	4	0.02	0.1	4	1.0
	Total cost	$48,754	$229	1,227	99	$492
4	Not-confirmed, modified surgery (N = 44)					
	Physician visits	209	4.80	7.6	28	7.0
	Hospital stays	47	1.10	0.6	39	1.2
	Number of surgeries	82	1.90	1.0	43	1.9
	Total cost	$125,009	$2,841	1,831	44	$2,810
5	Not-confirmed, surgery elected after intervening medical treatment (N = 34)					
	Physician visits	188	5.50	5.7	29	6.5
	Hospital stays	41	1.20	0.4	34	1.2
	Number of surgeries	45	1.30	0.7	34	1.3
	Total cost	$119,751	$3,522	3,197	34	$3,484
6	Not-confirmed, surgery elected without intervening medical treatment (N = 51)					
	Physician visits	128	2.50	3.6	26	4.6
	Hospital stays	51	1.00	0.4	49	1.0
	Number of surgeries	62	1.20	0.7	51	1.2
	Total cost	$132,024	$2,589	3,526	51	$2,559
7	Not-confirmed, consultant's recommendation followed† (N = 291)					
	Physician visits	949	3.30	7.0	150	6.2
	Hospital stays	100	0.30	0.6	84	1.2
	Number of surgeries	131	0.50	0.9	81	1.6
	Total cost	$293,514	$1,009	2,102	177	$1,643

* Related surgeries.

† This group consists of cohorts 3, 4, and 5.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Impact of a Mandatory Second-Opinion Program on Medicaid Surgery Rates

SUZANNE GRISEZ MARTIN, M.P.A.,[*] MICHAEL SHWARTZ, PH.D.[†]
BERNADETTE J. WHALEN, B.S.,[‡] DEBORAH D'ARPA, M.B.A.,[*]
GRETA M. LJUNG, PH.D.,[†] JOHN HOLDEN THORNE, B.A.,[*]
AND ANNE E. MCKUSICK, B.A.[*]

The effect of the Massachusetts second-opinion program on the volume of elective surgery in the Medicaid population was assessed using two approaches: a study of the program experience and surgery decisions of 2,501 program referrals, and an analysis of Medicaid surgery rates before and after program implementation. Nonconfirmation rates, which averaged 14.5 per cent, varied by procedure from 4 per cent for cholecystectomy to 26 per cent for disc surgery. The patient's surgery decision was related to the outcome of the second-opinion consultation: 85.5 per cent of the confirmed patients had the originally proposed operation, as compared with 31 per cent of the nonconfirmed patients. In the year after program implementation, the program was associated with a 20 per cent reduction in the volume of those procedures covered by the program. The greatest percentage declines were for hysterectomies, meniscectomies, hemorrhoidectomies and tonsillectomies/adenoidectomies. The decline in surgery rates was attributed both to a direct effect on patients referred to the program and to a sentinel effect whereby fewer operations were proposed. We conclude that the mandatory second-opinion program in Massachusetts saved Medicaid $3 to $4 for every dollar spent.

Statement of the Problem

BETWEEN 1971 AND 1977, the rate of surgery performed in the United States increased by 27.5 per cent. By 1977, one surgical procedure was being performed for every 100 persons.[1] Whether this increase represents improved access to care or greater utilization of surgical treatment relative to medical treatment is under debate. But there is no doubt that this growth has contributed to the nation's steadily increasing health expenditures.

Various public and private health insurers are attempting to control surgery expenditures by encouraging or requiring their enrolled populations to obtain a second opinion on the need for proposed elective operations. Advocates of this approach argue that second-opinion programs are able to identify operations that can be deferred or not performed without jeopardiz-

---

From the Massachusetts Health Research Institute,[*] the School of Management, Boston University,[†] and the Massachusetts Department of Public Welfare.[‡]

This study was supported by a grant from the Health Care Financing Administration, U.S. Department of Health and Human Services (grant #18-P-97037/1-01) and with funds from the Commonwealth of Massachusetts. The authors would like to thank Richard Potts, Carol Greenfield and Joseph Blackburn for their contribution and the Health Care Foundations of Massachusetts for their cooperation.

Address for reprints: Suzanne Grisez Martin, 134 A Petty Rd., RD #2, Cranbury, NJ 08512.

0025-7079/82/0100/0021/$01.75 © J. B. Lippincott Co.

ing the patient's health. They assume that these programs save money by lowering the amount of elective surgery, both by a direct program effect on patients who decide against the operation after a negative second opinion and by a sentinel effect whereby fewer operations are actually proposed.[2] Second-opinion programs are also described as consumer oriented, since they increase the patient's role in treatment decisions by providing an additional opportunity to explore the advantages and risks of a particular operation.

In 1976 the Massachusetts legislature mandated the establishment of a second-opinion program for Medicaid recipients. This paper presents the results of a study of this program, including its effects on the number of operations performed on the Medicaid population and the consequent cost savings to Medicaid.[3]

Program Description

The Consultation Program for Elective Surgery (CPES) requires that certain Massachusetts Medicaid recipients proposed for any of eight elective procedures participate in a second-opinion program as a precondition for Medicaid reimbursement for the surgery. After completing program requirements, patients may choose to have the originally proposed operation even if the second-opinion consultants do not confirm the need for surgery, and physician and hospital claims will be reimbursed according to the Medicaid fee schedule.

The Department of Public Welfare is responsible for the design and implementation of the program. Medicaid recipients without other health insurance coverage, primarily children and women under the age of 65, must participate in the program if proposed for one of the following eight procedures: tonsillectomy/adenoidectomy; hysterectomy; cholecystectomy; submucous resection/rhinoplasty; hemorrhoidectomy; meniscectomy; excision and stripping of varicose veins; and disc surgery/spinal fusion. The Department contracts with the state's five health care foundations to administer the program. Each foundation covers an area corresponding to one of the state's five Professional Standard Review Organization (PSRO) regions.

The surgeon proposing the operation refers the patient to the foundation. When a patient is required to have a consultation, the foundation arranges an appointment with a board-certified/eligible surgeon in the appropriate surgical specialty. The consultant examines the patient and either does or does not confirm the need for surgery. The foundation is responsible for notifying the proposing surgeon, the hospital and the patient of the consultant's decision. If the first consultant does not confirm the need for surgery, it must also inform the patient of the requirement to receive another consultation if he/she wants to have the operation. The identity of the proposing surgeon is concealed on any information forwarded to the consultant, and the proposing surgeon is not told the name of the consultant. The consultant is not allowed to perform the operation.

The program was phased in over a 14-month period by region, beginning in March 1977. There are two different program models. Three regions use a "hands-on" model in which persons referred to the program are examined by a foundation-assigned surgeon who assesses the need for the proposed operation. The other two regions use a "desk-audit" model in which proposed patients are screened by the foundation using preset criteria developed by area physicians in consultation with the Department of Public Welfare. The foundation notifies those patients not meeting the criteria and their proposing surgeons that a hands-on consultation is required if the patient is still interested in having the proposed operation.

Methodology

Two distinct approaches were used to examine the impact of the CPES on the number of operations performed on the Massachusetts Medicaid population. The first was a study of the surgery decisions of all 2,501 persons referred to the program in the Bay State (Boston Metropolitan area) and Central Massachusetts (Worcester area) regions during a one-year period (May 1, 1977 to April 30, 1978 in the former area and October 1, 1977 to September 30, 1978 in the latter). These two areas cover 65 per cent of the eligible population, and both employ the hands-on model. They were selected for special study because they implemented the program the earliest, in March and August 1977, respectively.

Program files maintained by both foundations were used as the data source for information on the demographic characteristics and program experience of patients referred to the CPES. Consultant reports were reviewed to determine both the reasons for nonconfirmations and whether any conditions were imposed on the confirmations. The Medicaid billing system and hospital records were used to track each referred patient from the date of referral to CPES through December 31, 1978, to determine whether the patient had the originally proposed CPES operation, no surgery or a procedure related to the originally proposed operation. Related surgeries included diagnostic procedures (such as arthroscopies performed on patients proposed for meniscectomies) and alternative surgical procedures (such as dilations and curettages performed on women proposed for hysterectomies). Fifty-nine per cent of the patients were followed for one year or more; 36 per cent for six to 12 months, and 5 per cent for three to five months.

Summary information on the disposition of patients referred to the CPES in the Charles River, Western and Southeastern Massachusetts regions was extracted from monthly reports submitted to the Department by each foundation.

The second approach analyzed monthly surgery rates for each CPES procedure in each of the state's five regions over a 44-month period spanning CPES implementation. Physician claims submitted to Medicaid were used to determine the number of operations. A total of 27,000 claims was reviewed to ensure that only one claim for any single operation was counted and that the patient receiving the operation was covered by CPES. Monthly counts of the number of individuals eligible for the program were obtained from the Department's monthly summary of expenditure reports.

For each procedure in each region, 44 months of surgery rates were calculated from May 1975 (the earliest point at which surgery codes were fully implemented in the Medicaid billing system) to December 1978 (determined by project deadline). Monthly rates were used to ensure significant data points to estimate pre- and post-CPES levels of surgery. Because of the 14-month spread in implementation dates among the five regions, the number of months of pre- and post-CPES data varied by region. There were between 22 and 37 months of pre-CPES data and between seven and 22 months of post-CPES data.

Monthly surgery rates for each procedure in each region were analyzed to determine if a statistically significant change ($p < 0.05$) in rates occurred between the pre- and the post-CPES periods.[3] When monthly surgery rates were independent, the statistical significance of a change in the level of surgery was performed using an ordinary t-test, or, when low surgery volume suggested nonnormality, a Mann-Whitney U test. When autocorrelation in the rates was present, the statistical significance of a change in level was tested using the procedure proposed by Box and

Tiao.[5] Basically, this amounts to performing an ordinary t-test after allowing for correlation among the observations.

The analysis was complicated by the fact that Medicaid reduced reimbursement fees for surgery by 30 per cent in February 1976. The same approach used to assess the impact of CPES was used to analyze the impact of the fee cut. When the fee cut had a statistically significant impact on monthly surgery rates, data prior to the fee cut were eliminated before evaluating the impact of CPES.

When a statistically significant change was detected in monthly surgery rates after CPES implementation, two rival hypotheses were considered as alternative explanations. The first hypothesis was that the change was an artifact of a gradual declining trend in surgery rates that had begun prior to the program and continued after the program came into effect. This possibility was evaluated by analyzing the statistical significance of a trend in the pre-CPES surgery rate data.

The second hypothesis was that factors unrelated to CPES influenced surgery rates at the time CPES was implemented. For Bay State and Central Massachusetts, this hypothesis was analyzed by considering as a nonequivalent control group the three regions of the state that had not yet implemented a CPES program. The monthly surgery rates of this control group were examined to determine if a statistically significant drop in rates occurred concurrently with CPES implementation in Bay State and Central Massachusetts. Such a drop in the control group would indicate that factors other than CPES might have caused the detected changes in Bay State and Central Massachusetts.

Estimates of the number of operations not performed as a result of CPES, that is, forgone surgeries, were determined from the analyses of program referrals and surgery rates. The number of forgone surgeries calculated from the analysis of program referrals provided an estimate of the direct impact of CPES on persons proposed for surgery and referred to the program. The number of forgone surgeries calculated from the surgery rate analysis provided an estimate of the total impact of the program, that is, the direct impact on persons referred to the program, as well as any impact that results because fewer people were initially proposed for surgery (the sentinel effect). To estimate the sentinel effect of CPES, the direct program impact based on program participants was subtracted from the total impact based on the surgery rate analysis.

The cost-effectiveness analysis involved a comparison of program expenditures and program savings to Medicaid that resulted from the forgone surgeries. Program expenditures included administrative costs of the foundations and the Department of Public Welfare, fees paid to physician consultants and estimated Medicaid payments for diagnostic tests ordered during consultations. Estimates of gross savings to Medicaid were calculated by multiplying estimates of the number of forgone surgeries by the average Medicaid payment for each procedure. Average payments per procedure were calculated from a review of the Medicaid claims of 372 program participants that had CPES operations and hospital per diem rates for each PSRO based on information from the Massachusetts Rate Setting Commission. Surgery-related costs included average Medicaid payments to the hospital, surgeon, assistant surgeon and anesthesiologist; payment for drugs and diagnostic tests ordered during hospitalization; and follow-up visits to the surgeon of record.

Gross program savings were adjusted to account for costs of alternative treatment for patients forgoing the CPES surgery. These costs were determined from an analysis of Medicaid claims for those program participants in Bay State and Central Massachusetts who did not have the originally proposed surgery. Alternative surgical costs were estimated from actual Medicaid payments for related surgeries of program participants. Alternative medical treatment costs for each procedure were estimated by comparing Medicaid payments for a sample of patients having surgery, less the surgery costs, with Medicaid payments for a sample of patients forgoing surgery. Medicaid payments for a six-month period after the date of program referral for all services possibly related to the surgery recommendation were calculated for each sample, and the geometric means (the means after logarithmic transformation of the data) were compared.

Findings

Program Statistics

Bay State and Central Massachusetts. Of the 2,501 persons referred to the Consultation Program for Elective Surgery (CPES) in these regions, most were proposed for tonsillectomies/adenoidectomies (40 per cent of all referrals) and hysterectomies (24 per cent of all referrals). Patients proposed for each of the other six CPES procedures accounted for no more than 10 per cent of all referrals. One third of the referred patients were under age 15, and 85 per cent of those over age 15 were women.

Eighty-one per cent of the referred patients had a physician consultation; 10 per cent canceled from the program prior to a consultation and 9 per cent had the consultation requirement waived (see Table 1). The majority of cancellations were patients who failed to keep a scheduled appointment. Other patients canceled because they were inappropriately referred (they had other insurance or were proposed for non-CPES procedures) or were admitted as emergencies (and thus not required to participate).

Waivers from the consultation requirement were granted when the nurse-coordinator at the foundation accepted the proposing surgeon's request for a waiver because of medical necessity (patient in severe pain or already in hospital) or when the patient was 15–20 miles from the nearest consultant (fewer than 10 per cent of all waivers were granted for travel hardship). Medical waivers were granted most frequently for patients proposed for disc surgery/spinal fusion and for hysterectomies, primarily for patients with the diagnosis of cancer in situ.

TABLE 1. Distribution of 2,501 Persons Referred to the CPES by Proposed Procedure: Bay State and Central Massachusetts

Procedure	Number Having Consultation (%)	Number Canceling (%)	Number Granted Waiver (%)	Total Number Referred (%)
Tonsillectomy/adenoidectomy	910 (89.9)	92 (9.1)	10 (1.0)	1,012 (100)
Hysterectomy	425 (70.3)	62 (10.2)	118 (19.5)	605 (100)
Cholecystectomy	202 (79.2)	23 (9.0)	30 (11.8)	255 (100)
Submucous resection/ rhinoplasty	172 (84.3)	30 (14.7)	2 (1.0)	204 (100)
Disc surgery/spinal fusion	84 (62.2)	14 (10.4)	37 (27.4)	135 (100)
Meniscectomy	109 (82.6)	12 (9.1)	11 (8.3)	132 (100)
Excision and stripping of varicose veins	89 (85.6)	9 (8.6)	6 (5.8)	104 (100)
Hemorrhoidectomy	43 (79.6)	8 (14.8)	3 (5.6)	54 (100)
Total	2,034 (81.3)	250 (10.0)	217 (8.7)	2,501 (100)

The final decisions of the physician consultants on the 2,034 patients who received consultations are presented in Table 2. For 85.5 per cent of the patients, the first consultant confirmed the need for the proposed surgery. The confirmation rate exceeded 90 per cent for patients proposed for cholecystectomy, excision and stripping of varicose veins, and hemorrhoidectomy. Review of the consultation reports indicated that 3 per cent of the confirmations were stated conditionally, for example, confirmed *if* the patient lost weight, or *if* an arthroscopy indicated the need.

The first consultant did not confirm the need for the proposed surgery for 294 patients. Almost half of these (44 per cent) had a second consultation. Of the nonconfirmed patients who sought a second consultation, 65 per cent were confirmed for CPES surgery.

The highest nonconfirmation rates were for patients proposed for disc surgery/spinal fusion (26 per cent), hysterectomy (23 per cent) and meniscectomy (24 per cent). The most frequent reason stated for nonconfirmation was that the proposed operation was "not medically necessary." Approximately 35 per cent of the nonconfirmations included qualifying statements by the consultant as to why the proposed procedure was not the optimum treatment at the time of the consultation. These reasons included recommendations that medical therapy or less extensive surgery be explored first; the need for more testing and/or a longer observation period; and statements that the patient was a poor risk for surgery.

Ten per cent of the patients received one or more diagnostic tests from the first consultant. The most frequently ordered tests were blood tests, Pap smears, anoscopies and cultures.

Charles River, Western and Southeastern Massachusetts. The program disposition of patients referred to the CPES in the Charles River, Western and Southeastern

TABLE 2. Distribution of 2,034 Persons Receiving Physician Consultations Through the CPES by Proposed Procedure and Outcome of the Consultation: Bay State and Central Massachusetts

	Distribution of Patients by Outcome of the First Consultation		Distribution of Patients Not Confirmed by the First Consultant		
Procedure	Confirmed by First Consultant No. (%)	Not Confirmed by First Consultant No. (%)	No Second Consultation No. (%)	Confirmed by Second Consultant No. (%)	Not Confirmed by Second Consultant No. (%)
Tonsillectomy/adenoidectomy	814 (89.5)	96 (10.5)	68 (70.8)	21 (21.9)	7 (7.3)
Hysterectomy	326 (76.7)	99 (23.3)	47 (47.5)	35 (35.3)	17 (17.2)
Cholecystectomy	193 (95.5)	9 (4.5)	4 (44.4)	4 (44.4)	1 (11.1)
Submacous resection/rhinoplasty	142 (82.6)	30 (17.4)	13 (43.3)	9 (30.0)	8 (26.7)
Disc Surgery/spinal fusion	62 (73.8)	22 (26.2)	7 (31.8)	10 (45.5)	5 (22.7)
Meniscectomy	83 (76.1)	26 (23.9)	18 (69.2)	3 (11.5)	5 (19.2)
Varicose veins	81 (91.0)	8 (9.0)	5 (62.5)	2 (25.0)	1 (12.5)
Hemorrhoidectomy	39 (90.7)	4 (9.3)	2 (50.0)	1 (25.0)	1 (25.0)
Total	1,740 (85.5)	294 (14.5)	164 (55.8)	85 (28.9)	45 (15.3)

TABLE 3. Distribution of 2,363 Patients Referred to the CPES: Charles River, Western Massachusetts and Southeastern Massachusetts

Region	No. Referred	No. (%) of Referrals Meeting Criteria	No. Having One Consultation	% Confirmation Rate: First Consultation	No. Having Two Consultations
Charles River	257	not applicable	222	87.8	13
Western Massachusetts	942	732 (77.2)	143	74.1	25
Southeastern Massachusetts	1,164	1,080 (92.8)	74	91.9	4

Massachusetts regions during the first year of operation in each site is presented in Table 3. The Charles River statistics are similar to those of Bay State and Central Massachusetts, but of smaller volume since the area encompasses only 3 per cent of the eligible CPES population. In Western and Southeastern Massachusetts, where the desk-audit model is used, 2,106 patients were screened using the preset criteria. Seventy-seven per cent of Western referrals and 93 per cent of Southeastern referrals met the criteria and were therefore not required to have a second-opinion consultation. The confirmation rate at the first consultation for those patients not meeting the criteria was 74 per cent in Western (n = 143) and 92 per cent (n = 74) in Southeastern Massachusetts. Only 29 patients in both areas had two second-opinion consultations.

Effect of the CPES on Number of Surgeries Performed

Surgery Decisions of Bay State and Central Massachusetts Patients. Seventy-two per cent of the 2,501 patients referred to the CPES had the originally proposed procedure; 22 per cent had no surgery, and 4 per cent had a related surgery. The

TABLE 4. Surgery Decisions of 2,501 Persons Referred to the CPES by Proposed Procedure: Bay State and Central Massachusetts

Procedure	No Surgery No. (%)	CPES Surgery No. (%)	Related Surgery No. (%)	Outcome Unknown No. (%)	Total No. (%)
Tonsillectomy/ adenoidectomy	211 (21.8)	786 (77.7)	2 (0.2)	3 (0.3)	1,012 (100.0)
Hysterectomy	118 (19.5)	423 (69.9)	53 (8.8)	11 (1.8)	605 (100.0)
Cholecystectomy	36 (14.1)	211 (82.8)	1 (0.4)	7 (2.7)	255 (100.0)
Submucous resection rhinoplasty	64 (31.4)	128 (62.7)	1 (0.5)	11 (5.4)	204 (100.0)
Disc surgery/ spinal fusion	31 (23.0)	84 (62.2)	17 (12.6)	3 (2.2)	135 (100.0)
Menisceetomy	34 (25.8)	67 (50.8)	25 (18.9)	6 (4.5)	132 (100.0)
Varicose veins	36 (34.6)	67 (64.4)	0 (0.0)	1 (1.0)	104 (100.0)
Hemorrhoidectomy	15 (27.7)	34 (63.0)	5 (9.3)	0 (0.0)	54 (100.0)
Total	555 (22.2)	1,800 (71.9)	104 (4.3)	42 (1.7)	2,501 (100.0)

surgery status of 2 per cent of the patients is unknown because they were ineligible for Medicaid at the time of follow-up, the Medicaid number was unavailable, or there was an unlisted prodecure code during a hospitalization. Ninety-two per cent of the patients having surgery did so within three months of the date on which their proposing surgeon referred them to the program. Patients proposed for cholecystectomy, tonsillectomy/adenoidectomy and hysterectomy were most likely to have the originally proposed operation (see Table 4). Related surgery as an alternative to the CPES surgery was most frequently performed on patients proposed for meniscectomy (mostly arthroscopies), disc surgery (mostly myelograms) and hysterectomy (mostly dilations and curettages).

There was a strong relationship between the first consultant's opinion and the patient's surgery decision: 85.5 per cent of patients whose need for a CPES operation was confirmed by the first consultant had the surgery, as compared with 31 per cent of those not confirmed (see Table 5). Of patients not confirmed by the first consultant, those least likely to have the CPES operation were those deciding against a second consultation (only 7 per cent of these patients had the originally proposed operation, in apparent disregard of CPES regulations). Forty per cent of the patients not confirmed by either the first or second consultant had the operation anyway, as did 74 per cent of those who received a nonconfirmation from the first consultant and a confirmation from the second consultant. Seventy-eight per cent of the patients waived from the consultation and 20 per cent of those patients who canceled also had the originally proposed operation.

To identify the possible range of program impact, project staff made both minimum and maximum estimates of the number of patients referred to the CPES who did *not* undergo the proposed operation as a result of the program. The minimum estimate assumed that only a portion of those patients who received nonconfirmations were deterred from surgery by the program. The number of nonconfirmed patients who would have decided against surgery without the pro-

TABLE 5. Distribution of 2,501 Persons Referred to the CPES by Program Disposition and Surgery Decision: Bay State and Central Massachusetts

Program Disposition	No Surgery No. (%)	CPES Surgery No. (%)	Related Surgery No. (%)	Unknown Surgery No. (%)	Total No. (%)
Confirmed first consultation	180 (10.3)	1,487 (85.5)	53 (3.0)	21 (1.2)	1,740 (100.0)
Not confirmed, first consultation, no second consultation	137 (83.5)	11 (6.7)	12 (7.3)	4 (2.4)	164 (100.0)
Not confirmed first consultation, confirmed second	15 (17.6)	63 (74.1)	5 (5.9)	2 (2.4)	85 (100.0)
Not confirmed first consultation, not confirmed second	23 (51.1)	18 (40.0)	2 (4.4)	2 (4.4)	45 (100.0)
Granted waivers	24 (11.1)	170 (78.3)	17 (7.8)	6 (2.8)	217 (100.0)
Canceled	176 (70.4)	51 (20.4)	16 (70.4)	7 (2.8)	250 (100.0)
Total	555 (22.2)	1,800 (72.0)	104 (4.1)	42 (1.7)	2,501 (100.0)

gram was estimated from the percentage of confirmed patients who elected not to have surgery. This number was subtracted from the number of nonconfirmed patients who did not have CPES operations to establish the minimum estimate. The maximum estimate assumed that the patients who decided against surgery because of the program included both those who canceled from the program by not appearing for their consultation and those who had a consultation and did not have the surgery. These patients might have decided against surgery because of information provided by the consultant or due to the additional time the program gave them to consider the operation. CPES operations might not be performed on patients canceling without completing program requirements because provider claims would be subject to disallowance by Medicaid.

The minimum estimate of program impact on program participants in Bay State and Central Massachusetts is 177 forgone surgeries, representing 27 per cent of all patients referred to the program in these regions who did not have the originally proposed surgery. The maximum estimate of forgone surgeries is 577, or 85 per cent of all program participants not having the originally proposed surgery in Bay State and Central Massachusetts. This maximum estimate includes 232 confirmed patients, 194 nonconfirmed patients and 131 patients who failed to keep scheduled appointments. The number of forgone surgeries by procedure from this analysis of program participants is found in Table 9.

CPES Impact on Medicaid Surgery Rates

Table 6 presents the mean monthly surgery rates per 100,000 population by region for each procedure for the periods before and after CPES implementation. The table also indicates whether the change in surgery rates after CPES implementation is statistically significant. (In the Appendix, several graphs of the time series in Bay State and Central Massachusetts are presented and discussed).

Minimum and maximum estimates were made of the magnitude of CPES impact on Medicaid surgery rates. The minimum estimate was based only on data from the Bay State and Central Massachusetts regions. At least 15 months of post-CPES surgery rates were available for both of these regions, and the possibility that other factors caused the surgery rate decline after CPES implementation could be considered by comparison with the nonequivalent control group.

Three criteria were used to assess the effect of CPES on surgery rates in Bay State and Central Massachusetts: 1) whether there was a statistically significant drop in the mean rate of surgery after CPES implementation; 2) whether there was evidence of a statistically significant decline in the rate of surgery prior to CPES (thus suggesting that the rate drop might be an artifact of a declining trend); and 3) whether there was a comparable drop in mean surgery rates in the control group (suggesting that factors other than CPES were responsible for the observed decline). Procedures that met all three criteria were included in the minimum estimate of CPES impact.

The maximum estimate was composed of all procedures in any region in which a statistically significant drop in the mean level of surgery rates occurred concurrently with program implementation and in which no evidence of a trend prior to CPES was found. This statewide estimate should be considered preliminary since less than 12 months of post-CPES surgery rates were available for three of five regions.

The minimum estimate attributes 603 fewer surgeries in the Bay State and Central Massachusetts regions to the CPES in the year after program implementation, including 391 fewer tonsillectomies/adenoidectomies and 135 fewer hysterec-

TABLE 6. Average Monthly Surgery Rates Per 100,000 Medicaid Population in the Pre-CPES and Post-CPES Period By Procedure and Region*

Procedure	Bay State		Central		Western		Southeastern		Charles River	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Tonsillectomy/adenoidectomy	30.77	18.25†	64.11	48.97†	99.98	49.04†	86.84	49.25†	75.83	55.69
Hysterectomy	17.10	13.10†	15.41	11.73†	28.61	19.67†	22.98	16.51†	19.70	9.49†
Cholecystectomy	9.01	9.07	15.79	13.20	12.16	12.14	15.85	10.37†	8.93	16.82
Submucous resection/rhinoplasty	2.65	2.34	13.68	7.44†	9.34	5.74†	16.28	15.43	12.63	10.58
Disc surgery/spinal fusion	4.35	4.14	6.80	5.64	5.55	5.46	4.68	4.13	5.25	5.26
Meniscectomy	3.32	2.46†	2.82	3.86	2.13	1.95	3.88	2.13†	2.70	5.22
Varicose veins	3.32	2.39†	4.23	3.39	4.65	3.23	3.30	1.70	4.88	3.15
Hemorrhoidectomy	2.76	2.11†	5.10	3.40†	4.51	3.32	3.38	1.90†	3.97	2.15

* Pre-fee cut data were eliminated for those cases in which fee cut had a statistically significant effect on monthly surgery rates. The Pre- and Post-CPES periods in each region are as follows:

Bay State	1–22, 23–44
Central	1–29, 30–44
Western	1–34, 35–44
Southeastern	1–37, 38–44
Charles River	1–37, 38–44

(Month 1 corresponds to May 1975; month 44 corresponds to December, 1978.)

† Indicates that change in level is statistically significant ($p < 0.05$).

tomies (see Table 7). These 603 operations represent a 20 per cent decline in total CPES operations in these two regions, as compared with the year before implementation. The greatest percentage declines were for tonsillectomies/adenoidectomies (28 per cent), hysterectomies (23 per cent) and meniscetomies (22 per cent).

The maximum estimate of CPES impact attributes 1,139 forgone surgeries to CPES statewide, representing a 20 per cent decline in CPES operations statewide in the year after implementation, as compared with the preceding year. Hysterectomies showed the greatest decline (26 per cent), followed by hemorrhoidectomies (23 per cent), meniscectomies (23 per cent) and tonsillectomies/adenoidectomies (22 per cent). The only procedure with no evidence of a statistically significant decline in surgery rates in any region was disc surgery/spinal fusion. The percentage decline in CPES operations varied by region from a 7–8 per cent decline in two regions to a greater than 20 per cent decline in three.

The Sentinel Effect

The sentinel effect, which refers to the effect of the program on the amount of proposed surgery, was measured by subtracting the direct program effect (estimated from the analysis of program participants) from the total program effect (estimated from the surgery rate analysis). In Bay State and Central Massachusetts, for which there are minimum and maximum estimates of both the direct and total impact, there are corresponding minimum and maximum estimates of the sentinel effect. The minimum estimate of the sentinel effect was calculated by subtracting the maximum estimate of direct impact from the minimum estimate of total impact; the maximum estimate of the sentinel effect was calculated by subtracting the minimum estimate of direct impact from the maximum estimate of total impact. Table 8 presents the minimum and maximum estimates of the sentinel effect for Bay State and Central Massachusetts. Tonsillectomies/adenoidectomies show

TABLE 7. Minimum and Maximum Estimates of the Number of Forgone CPES Operations Attributable to the CPES by Procedure Based on Surgery Rate Analysis

Procedure	Minimum Estimate (Bay State and Central)			Maximum Estimate (Statewide)		
	No. Forgone Operations	No. Operations 12 Months Prior to CPES	% Change	No. Forgone Operations	No. Operations 12 Months Prior to CPES	% Change
Tonsillectomy/ adenoidectomy	391	1,385	28	641	2,878	22
Hysterectomy	135	595	23	276	1,054	26
Cholecystectomy	0	374	0	45	622	7
Submucous resection/ rhinoplasty	33	170	19	71	429	17
Disc surgery/spinal fusion	0	138	0	0	254	0
Meniscectomy	25	114	22	39	172	23
Varicose veins	0	126	0	27	196	14
Hemorrhoidectomy	19	113	17	40	175	23
Total	603	3,015	20	1,139	5,780	20

the most evidence of a sentinel effect. There is some sentinel effect for all procedures except cholecystectomies and disc surgery, but, with the exception of hysterectomies, the number of operations involved is small.

In Western and Southeastern Massachusetts, where the desk-audit model is used, the maximum direct program effect was estimated as the number of cases that did not meet the preset criteria. A minimum sentinel effect was calculated by subtracting this maximum estimate of direct effect from the total effect based on the estimates from the surgery rate analysis. In Southeastern Massachusetts the sentinel effect seems to account for most of the impact of CPES on tonsillectomies/adenoidectomies, cholecystectomies and hysterectomies, since very few patients did not meet the criteria and fewer still received nonconfirming second opinions. In Western Massachusetts there was evidence of a sentinel effect only for submucous resection.

Projected Savings of CPES

Medicaid expenditures for CPES during the first year of program operation in all regions were $293,800, with 77 per cent of the budget spent for administration, 22.5 per cent for physician fees and 0.5 per cent for diagnostic tests ordered by consultants and billed to Medicaid.

Costs of alternative treatment for those patients forgoing CPES operations were grouped into costs of alternative surgical treatment and costs of alternative medical treatment. Costs for alternative surgical treatment ranged by procedure from 0 per cent of unadjusted gross savings for varicose veins to 40 per cent for hemorrhoidectomy. On average, the cost of alternative surgical treatment represented 8 per cent of gross savings.

No adjustment was made for alternative medical treatment costs because there was no statistically significant difference between the average Medicaid payments for patients referred to CPES who had the

TABLE 8. Minimum and Maximum Estimates of the Number of Forgone Operations Attributed to the Sentinel Effect by Region and Procedure

	Bay State		Central Massachusetts		Western Massachusetts	Southeastern Massachusetts
Procedure	Minimum No. (%)*	Maximum No. (%)*	Minimum No. (%)*	Maximum No. (%)*	Minimum No. (%)*	Minimum No. (%)*
Tonsillectomy/adenoidectomy	231 (59)	350 (90)	0	55 (67)	—	166 (99)
Hemorrhoidectomy	3 (16)	18 (95)	0	9 (100)	—	0
Cholecystectomy	—	—	—	—	—	42 (93)
Menisceectomy	0	8 (32)	—	—	—	9 (64)
Submucous resection/rhinoplasty	0	0	7 (21)	26 (79)	24 (83)	—
Varicose veins	0	23 (85)	—	—	—	—
Hysterectomy	2 (2)	72 (63)	6 (30)	12 (60)	0	15 (25)
Disc surgery/spinal fusion	—	—	—	—	—	—
Total	236 (40)	471 (80)	13 (9)	102 (71)	24 (24)	230 (79)

The symbol — indicates that CPES had no statistically significant impact on surgery rates; 0 indicates that CPES had an impact, but that none of the impact is attributed to the sentinel effect.

* Numbers in parentheses denote the percentage of the total program impact that is due to the sentinel effect.

CPES operation (less payment for the surgery) and for patients forgoing the CPES operation.

The statewide estimate of net program savings for the year following CPES implementation in all regions is $856,506, for a 1:3.9 cost–benefit ratio. This estimate was derived from the 1,139 forgone CPES procedures statewide in the year after program initiation.

There are four estimates of program savings for the Bay State and Central Massachusetts regions, each derived from one of the four different estimates of forgone surgeries. The number of forgone surgeries and corresponding net program savings by procedure for each of these four estimates for the two regions is presented in Table 9. Savings estimates for these two regions range from $96,184, for a cost–benefit ratio of 1:1.5, to $657,265, for a cost–benefit ratio of 1:4.7.

Discussion

Gertman et al. have reported that the CPES "may produce only modest direct savings in Medicaid expenditures."[6] This conclusion was based on an analysis of the program disposition of CPES referrals to the Bay State Health Care Foundation and a follow-up of those patients referred for hysterectomies. Program benefits were calculated by comparing the estimated costs associated with surgery events of women who received negative second opinions with the estimated costs for these women if their surgery experience paralleled that of women with positive opinions. The authors estimate a 2.27 benefit–cost ratio for hysterectomies in Bay State. Using similar assumptions in a "more speculative" analysis, for which the surgery decisions of all referred patients were not known, they estimate an overall benefit–cost ratio of 1.1 for CPES in Bay State.

Our minimum estimate of program savings based on program referrals in Bay State, which also assumes a CPES effect only on some patients not confirmed for the proposed surgery, also predicts somewhat modest savings to Medicaid (a benefit–cost ratio of 1.83, based on follow-up of all Bay State referrals and including costs of alternative treatment). However, other aspects of our analysis have lead us to conclude that this estimate is unrealistically low because it excludes any sentinel effect and any direct impact on referrals other than those with negative opinions on the need for the proposed operation.

The only other published data on mandatory second-opinion programs come from several security welfare funds in the greater New York City area. McCarthy has analyzed these programs through observation of program participants and has reported on referral volume, consultation outcomes and surgery status. In 1978, McCarthy reported an overall nonconfirmation rate of 18.5 per cent on the 5,025 second-opinion consultations provided in the mandatory program since 1972.[7] This is slightly higher than the 14.5 per cent nonconfirmation rate for 2,501 consultations in the CPES during its first year in Bay State and Central Massachusetts. In both programs approximately two thirds of the patients not confirmed did not have the originally proposed surgery. McCarthy reports that 79.4 per cent of the nonconfirmed/no surgery patients interviewed stated that they had elected against surgery because the consultant had advised against it.

The assessment of the Massachusetts second-opinion program presented here includes both an analysis of program participants and an analysis of the target population's surgery rates before and after program implementation. These two approaches result in different estimates of program impact. The advantage of the analysis of program participants is that the number of forgone surgeries is based on a direct count of persons referred to CPES. Its disadvantages are the difficulty of de-

TABLE 9. Minimum and Maximum Estimates of Net Program Savings* in the Bay State and Central Massachusetts Regions for the Year Following Program Implementation

	Estimated Net Savings Based on Analysis of Program Participants				Estimated Net Savings Based on Surgery Rate Analysis			
	Minimum		Maximum		Minimum		Maximum	
Procedure	No. Forgone Operations	Net Savings ($)	No. Forgone Operations	Net Savings ($)	No. Forgone Operations	Net Savings ($)	No. Forgone Operations	Net Savings ($)
Tonsillectomy/ adenoidectomy	68	−37,973	206	38,645	391	153,420	473	196,945
Hysterectomy	51	69,660	127	249,769	135	265,633	135	265,633
Cholecystectomy	4	−4,761	32	62,311	0	−15,805	0	−15,805
Submucous resection/ rhinoplasty	17	3,853	59	45,972	33	13,289	42	23,874
Disc surgery/ spinal fusion	13	44,390	35	141,520	0	−10,093	0	−10,093
Meniscectomy	18	21,337	51	64,670	25	27,834	25	27,834
Varicose veins	5	1,496	30	40,078	0	−6,281	27	37,301
Hemorrhoidectomy	1	−1,818	17	14,300	19	15,248	28	22,187
Total	177	96,184	557	657,265	603	443,245	730	547,876
Cost–benefit ratio†		1:1.5		1:4.7		1:3.5		1:4.1

* Net program savings equal gross savings less estimated payments for alternative treatment and program costs.

† Cost–benefit ratio equals gross savings less estimated payments for alternative treatment divided by program costs. For Bay State and Central Massachusetts, the program costs were $178,747.

termining which program participants decided against surgery because of the program and the impossibility of measuring the sentinel effect (fewer operations proposed because of the program).

The difficulty of assessing the impact of a second-opinion program based solely on simple confirmation or nonconfirmation rates was highlighted in Paris' report on the voluntary second-opinion program for New York City's employees.[8] Paris reported that patients often received much more information than a mere confirmation/nonconfirmation on the need for the proposed operation from the surgeon rendering the second opinion. In a review of consultant reports he found a large "gray area" involving 22 per cent of the cases, in which the consultant supplemented the negative/positive opinion with "qualifications." These "qualifications" included such statements as: "observe patient further." and "additional diagnostic information or evaluation on the need for the proposed surgery required." The statement written in the records probably reflects more extensive conversations between the consultant and the patient on the advantages and disadvantages of various treatment options. This additional communication may explain why some nonconfirmed cases decided to have the originally proposed procedure and why some confirmed patients decided not to. For this reason, nonconfirmation rates are of limited value in predicting which program participants are affected by the program.

The advantage of the analysis of surgery rates we have performed is that it focuses on the entire population eligible to participate in CPES and can thus measure both direct program impact and the sentinel effect without making any assumptions about which program participants were affected by the program. Furthermore, by taking into consideration pre- as well as post-CPES rates, two possible errors in the analysis are avoided: 1) attributing a decline in rates to CPES when the decline represents a continuation of preprogram trends and 2) interpreting a constant level of surgery after implementation as indicating no program effect on the surgery level, when it in fact represents a statistically significant decline from the pre-CPES baseline period.

The surgery rate analysis in Bay State and Central Massachusetts was strengthened by using a nonequivalent control group to evaluate whether factors other than CPES may have been responsible for any observed drop in surgery rates. The use of nonequivalent control groups to improve interrupted time series designs has been strongly recommended and applied in a variety of situations.[4,9,10] However, one should attempt to identify sources of bias that might result from nonequivalency and lack of randomization. An important source of nonequivalency in this case was surgery rates that were initially higher in the control area than in Central Massachusetts, and particularly in Bay State. Because of regression artifacts that might tend to lower surgery rates in the control area, there is a bias toward interpreting similar declines in Bay State and Central Massachusetts to something other than CPES, when in fact they may have been caused by CPES. Despite this potential bias, with one exception, there was no drop in surgery rates in the control area concurrent with CPES implementation in Bay State and Central Massachusetts.

One factor that has been cited as an alternative explanation for the decline in surgery rates is the PSRO program. As the PSRO program was at least as completely implemented in the control area as in Bay State and Central Massachusetts when CPES was implemented, if the PSRO program caused the decline in these regions, one would expect to observe a similar decline in the control (assuming that the

PSRO impact on surgery rates is similar statewide). However, as stated earlier, in only one case out of 16 possibilities did surgery rates in the control area decline when CPES was implemented in Bay State or Central Massachusetts.

A further illustration of the unlikelihood of the PSRO program causing the surgery rate decline that we have ascribed to CPES can be seen in Southeastern Massachusetts, where most of the CPES impact has been attributed to the sentinel effect. As Figure 1 shows, there was a drop in surgery rates for all procedures combined in month 15 (June 1976) and then a leveling off until month 37 (May 1978), when the CPES was implemented. The first drop *might* be accounted for by the PSRO program, which was implemented between April 1976 and July 1977, or by the February 1976 Medicaid fee cut. There is *no* reason to suspect, however, that the PSRO program caused the abrupt drop in rates following CPES implementation, when the PSRO program had been in effect for 10 months. Furthermore, the PSRO program is more likely to have a gradual than an abrupt impact.[11]

The surgery rate analysis is better suited for the detection of abrupt, rather than gradual, changes. Graphs of the time series indicate that when CPES did affect surgery rates, there was an immediate drop in the level of surgery after program implementation. In no case did post-CPES surgery rates exhibit a tendency to return to pre-CPES levels. However, post-CPES surgery rates were tracked for less than two years in all areas of the state. Thus, the longer-term impact of the program on surgery volume cannot be assessed by the project's data.

The approach used to assess the impact of the fee cut on surgery rates was to eliminate the pre-fee cut data only if a statistically significant impact was observed. However, since the impact of the fee cut might not have been detected because of the low power of the statistical tests, and since this approach might thus bias results, the analysis was performed with pre-fee cut data eliminated in all cases. There was only a minimal effect on our conclusions:

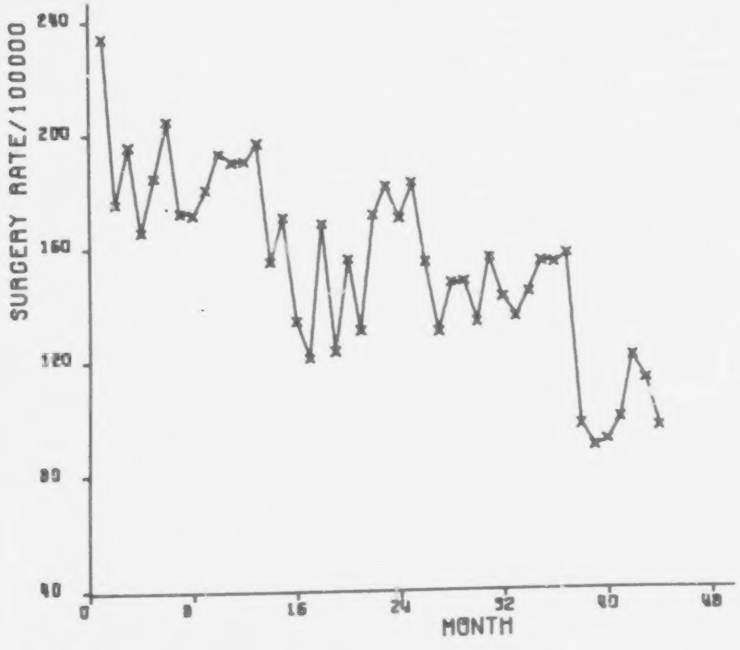


FIG. 1. Monthly rate of all CPES procedures in Southern Massachusetts.

the minimum cost–benefit ratio declined from 1:3.5 to 1:3.4; the maximum from 1:4.1 to 1:4.0; estimated statewide savings declined from $858,506 to $853,323.

Several points relative to the validity of the surgery rate data should be mentioned. The denominator data are based on monthly figures from routine Department of Public Welfare reports, and while there were monthly fluctuations, there were no significant differences in the age, sex or category-of-assistance distribution of the population over the 1976–1978 period or across the PSRO regions.

Numerator data are based on the physician claims submitted to Medicaid. This might be expected to result in some undercounting of the number of operations performed if there were no claims submitted for operations done by residents and if some physicians did not bill Medicaid because of low fees. In tracking which program participants had surgery, project staff compared hospital records with Medicaid files and found only 77 cases out of 1,800 (less than 5 per cent) in which a CPES operation was done without a corresponding physician bill in the Medicaid files. There is, therefore, a minimal amount of undercounting in the number of operations. However, only if a change in undercounting occurred once the CPES began would there be a bias in the analysis of surgery rates that has been presented.

There are two limitations of the surgery rate analysis. First, the approach may underestimate program impact for some procedures because the variability in monthly surgery rates made it unlikely that changes of less than 20 per cent would be identified as statistically significant. Second, the approach may overestimate the magnitude of CPES impact for other procedures, because if a statistically significant drop in surgery rates was detected, the entire amount of the drop was attributed to CPES, when other factors might have accounted for some of the decline.

The combination of both approaches provides insight into the effect of the CPES on different procedures. For example, despite the fact that there was no statistically significant drop in the rate of disc surgeries/spinal fusions in any region, the high nonconfirmation rate for this procedure suggests a possible program impact that was too small to be detected by the surgery rate analysis. Because of the expense of this procedure, a small impact on surgery rates would generate large dollar savings. For hysterectomies and meniscectomies, there were both a high rate of nonconfirmation and a large impact of the program on surgery rates. Tonsillectomies/adenoidectomies and hemorrhoidectomies had only a moderate nonconfirmation rate, but there was a large impact on surgery rates. This demonstrates that the nonconfirmation rate alone, which indicates only direct program effect, may not be a good indicator of the total impact of the program.

The effect of the program on surgery rates varied by region. Two of the regions using the hands-on model had a decline in surgery rates of over 20 per cent. The remaining region had a decline of slightly less than 10 per cent. However, in this region, which includes only 3 per cent of the eligible population, because of a short follow-up period, small volume and large variability in surgery rates, it was difficult to assess accurately the impact of the CPES. The short post-CPES period in the two regions with the desk-audit model permitted only a preliminary assessment of this model. One of these regions had a decline of more than 20 per cent in surgery rates following CPES; the other had a decline of just under 10 per cent. Since few people actually received consultations under the desk-audit model, program impact in these regions was attributed to a sentinel effect. It is interesting that the effect of the program on surgery rates did not seem related to the pre-CPES level of surgery. The area that had the lowest pre-CPES surgery rates had the largest decline (Bay State); the area with the second highest pre-CPES surgery rates had the second largest decline (Southeastern Massachusetts).

It is difficult to interpret the effect of CPES on surgery rates for each procedure statewide because of the multiple interventions: the fee cut plus the four different times at which CPES was phased in. However, it is interesting to compare the percentage decline in Medicaid surgery rates statewide by procedure with declines over the same period in surgery rates in the Northeastern region. As highlighted in Table 10, with the exception of disc surgery/spinal fusion, the decline in surgery rates in the Massachusetts Medicaid population was much larger than that in the Northeastern region.

TABLE 10. Yearly Percentage Change in Northeast Regional Surgery Rates (1975–1978)* Compared With Change in Massachusetts Medicaid Surgery Rates

Procedure	Regional Change (%)†	Massachusetts Medicaid Change (%)†
Tonsillectomy/ adenoidectomy	−11	−25
Hysterectomy	−8	−16
Cholecystectomy	−1	−7
Submucous resection/ rhinoplasty	+1	−8
Disc surgery/ spinal fusion	−3	−1
Meniscectomy	−3	−12
Varicose veins	−5	−19
Hemorrhoidectomy	−16	−26

* From Vital and Health Statistics, Data from the National Health Survey, Series 13, no. 34 (Surgical Operations in Short Stay Hospitals: United States, 1975); no. 37 (Utilization of Short Stay Hospitals, Annual Summary for the U.S., 1977); no. 46 (Utilization of Short Stay Hospitals, Annual Summary for the U.S., 1978); and personal communication, National Center for Health Statistics.

† The regional change was calculated from the coefficient of the line fit by ordinary least squares to the logarithm of the four yearly rates from 1975–1978; the Massachusetts change was calculated from the coefficient of the line fit by ordinary least squares to the logarithm of the monthly surgery rates in the post-fee cut period.

The minimum and maximum estimates based on the two approaches can be used to focus attention on the effect of different assumptions on conclusions regarding CPES impact. Estimates of net CPES savings in Bay State and Central Massachusetts range from $96,184 to $657,256. Project staff believe that the lowest estimate, $96,184, underrepresents CPES impact because it omits any sentinel effect, any effect on patients who do not have surgery after missing a scheduled appointment and who may be deterred by bureaucratic barriers, and any effect on patients confirmed for the proposed operation who may defer or forgo surgery because of information provided by the consultant or as a result of additional time to consider the decision. The maximum estimate, $657,256, which attributes all decisions against surgery of confirmed and canceling patients to the program, but also omits any sentinel effect, may overstate CPES impact. Actual program savings for these two areas are probably between $443,245 and $547,876. The $443,245 figure translates into a cost–benefit ratio of 1:3.5—a realistic estimate of CPES savings for the two regions with the most available data for this analysis.

While our savings estimates consider payments for alternative medical treatments, the approach used is admittedly rough because there was no way to identify which Medicaid claims were related to a forgone CPES procedure. If alternative medical treatment costs were small, they would not have been detected; however, they would therefore not alter the net savings figures reported to any great extent. While a longer follow-up on medical treatment costs would have been interesting, six months is reasonable, given McCarthy's finding that there was no increase in the number of patients having alternative medical treatment between the six- and 12- month follow-up periods.

It should be reemphasized that reported savings are only from the perspective of Medicaid and do not address any costs or benefits to patients, providers, other insurers or to society. Further, the project did not address important quality-of-care issues associated with having or forgoing any of the CPES procedures.

Conclusions

The Massachusetts second-opinion program was associated with a 20 per cent reduction in the volume of the elective procedures covered by the program. The program was associated with a statistically significant decline in surgery rates for seven of the eight procedures in at least one of the five PSRO regions. The greatest declines were for hysterectomies, hemorrhoidectomies, meniscectomies and tonsillectomies/adenoidectomies. The only procedure with no significant decline was disc surgery/spinal fusion.

The reduction in the level of surgery generated savings to the Medicaid program that resulted in a return of $3–$4 for every dollar spent. Thus, the Massachusetts Medicaid program has realized substantial savings from its second-opinion program.

Future assessments of second-opinion programs should attempt to examine the effect of second opinions on quality of care, as well as the programs' value in providing additional information to patients to improve their ability to evaluate proposed operations.

References

1. American College of Surgeons. Socioeconomic factbook for surgery 1979. Haug JN, Kuntzman K (eds). Chicago: American College of Surgeons, 1979:35.

2. Grafe WR, McSherry CK, Finkel ML, et al. The elective surgery second opinion program. Ann Surg 1978;188:323.

3. Martin SG, Shwartz M, Cooper DD, et al. The effect of a mandatory second opinion program on medicaid surgery rates: an analysis of the Massachusetts Consultation Program for Elective Surgery. Baltimore: Health Care Financing Administration, DHHS, 1980.

4. Glass GV, Willson VL, Gottman JM. Design and analysis of time series experiments. Boulder: Colorado University Press, 1975.

5. Box GEP, Tiao GC. A change in level of a non-stationary time series. Biometrika 1965;52:181.

6. Gertman PM, Stackpole DA, Levenson DK, et al. Second opinions for elective surgery. N Engl J Med 1980;302:1169.

7. McCarthy E, Finkel ML. Second opinion elective surgery programs: outcome status over time. Med Care 1978;16:984.

8. Paris M, Salsberg E, Berenson L. An analysis of nonconfirmation rates. JAMA 1979;242:2424.

9. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company, 1969.

10. Campbell DT. Reforms as experiments. Am Psychol 1969;24:409.

11. Brooke RH, Williams KN, Rolph JE. Controlling the use and cost of medical services: the New Mexico Experimental Medical Care Review Organization/a four year case study. Med Care 1980;16(suppl):9.

Appendix

In this section we illustrate our interpretation of changes in surgery rates attributable to CPES with graphs of the time series for hysterectomies and tonsillectomies/adenoidectomies in Bay State, Central Massachusetts and the control area (i.e., rates for Western, Charles River and Southeastern Massachusetts combined for the period before CPES was implemented in these areas). For each time series, the following issues were considered:

> Did the fee cut (month 9) cause a statistically significant change in surgery rates? If so, pre-fee cut data are eliminated to de-

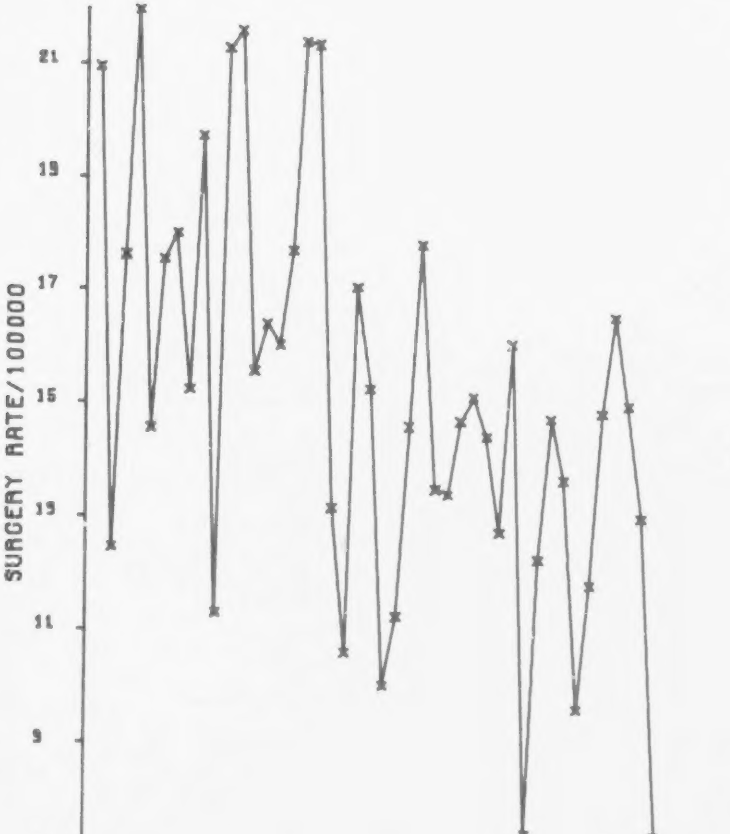


FIG. A-1. Monthly rate of hysterectomies in Bay State.

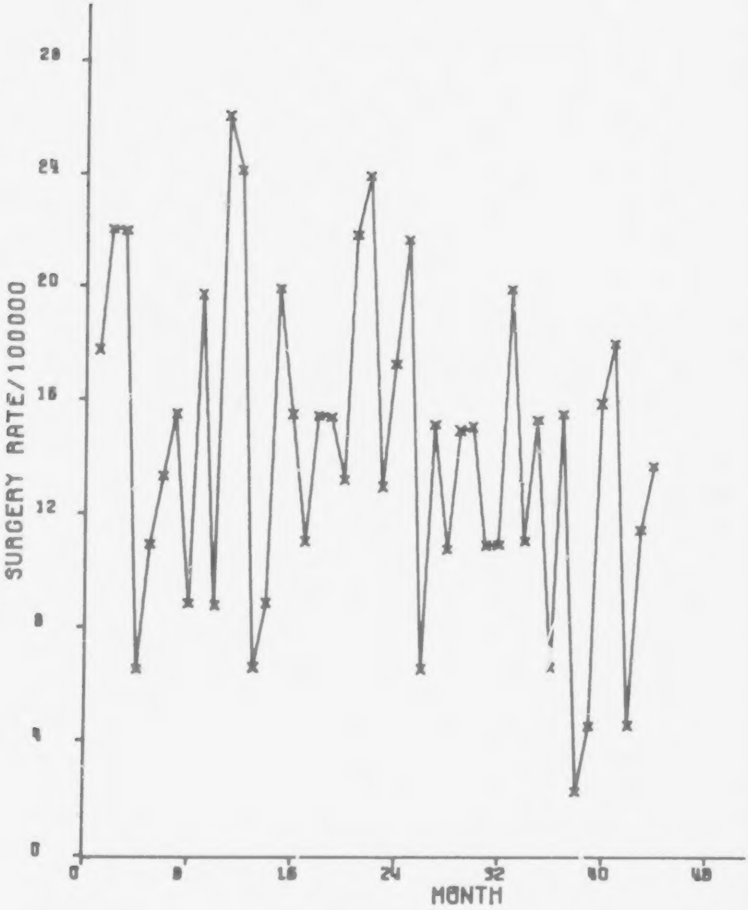


FIG. A-2. Monthly rate of hysterectomies in Central Massachusetts.

termine the pre-CPES mean level of surgery.

Was there a statistically significant decline in the mean level of surgery following CPES implementation (month 22 in Bay State, month 29 in Central Massachusetts)? The p value for the null hypothesis that the pre-CPES mean level of surgery equaled the post-CPES mean level is reported in the discussion.

Was there a statistically significant decline in the mean level of surgery in the control area following CPES implementation in Bay State or Central Massachusetts? If so, this would suggest that some factor other than CPES may have accounted for any drop in rates in Bay State or Central Massachusetts.

Should an observed drop in surgery rates after CPES implementation be attributed to the CPES or to a general decline in surgery rates over the entire time period? To evaluate this question, we explore evidence of a declining trend in pre-CPES rates in Bay State and Central Massachusetts and in the control. If there is evidence of a trend in rates prior to CPES, we considered a model to incorporate both a trend and a program impact (i.e., a linear regression model with a dummy variable).

If all evidence supports the hypothesis that CPES affected surgery rates, we include

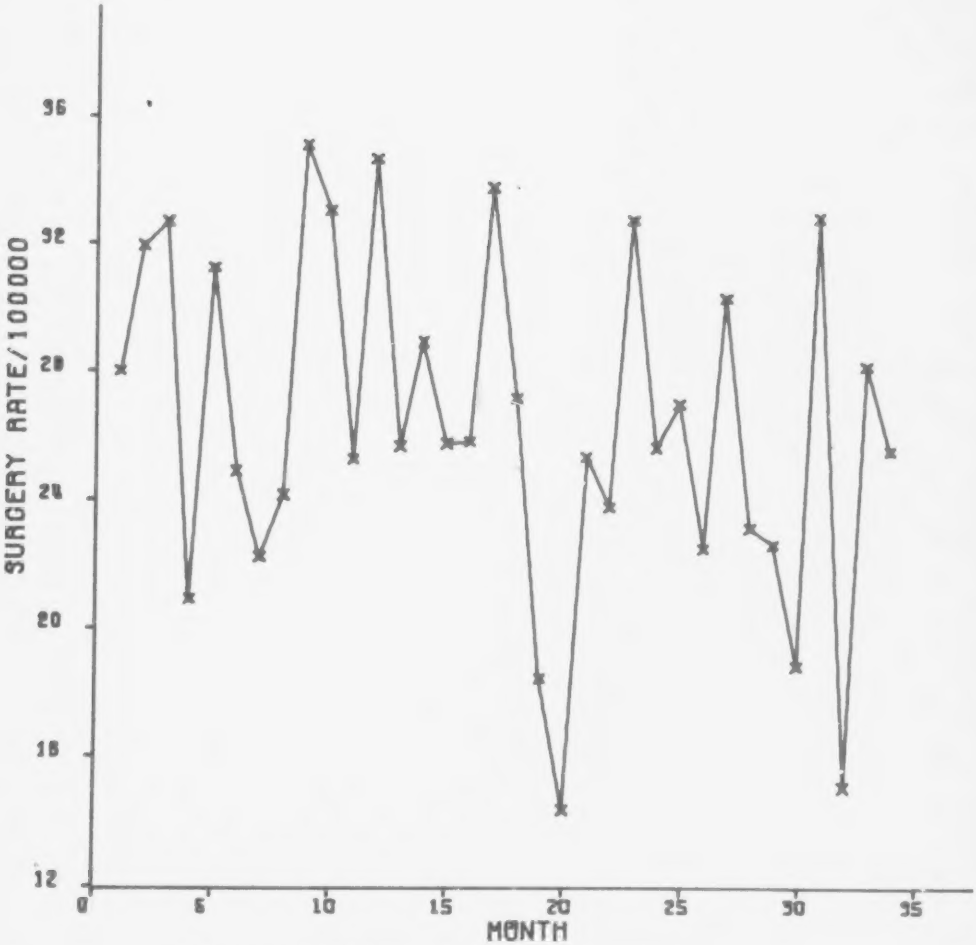


FIG. A-3. Monthly rate of hysterectomies in control area.

the effect in the minimum estimate of CPES impact. If there is a statistically significant drop in surgery rates following CPES implementation, but there is evidence to suggest that the drop may have been due to other factors, we include the effect only in the maximum estimate.

Procedures in Western, Charles River and Southeastern were considered only as candidates for inclusion in the maximum estimates of CPES impact since the post-CPES period was so short and there was no control area. The criterion for inclusion in these areas was that the change in the mean level of surgery following CPES implementaion was statistically significant.

Hysterectomies: Bay State (Fig. A-1) and Central Massachusetts (Fig. A-2)

In neither of these cases did the fee cut affect surgery rates. The mean level of surgery rates in the post-CPES period in both areas is lower than the mean level in the pre-CPES period ($p = 0.00$ in Bay State and $p = 0.02$ in Central). The drop in the

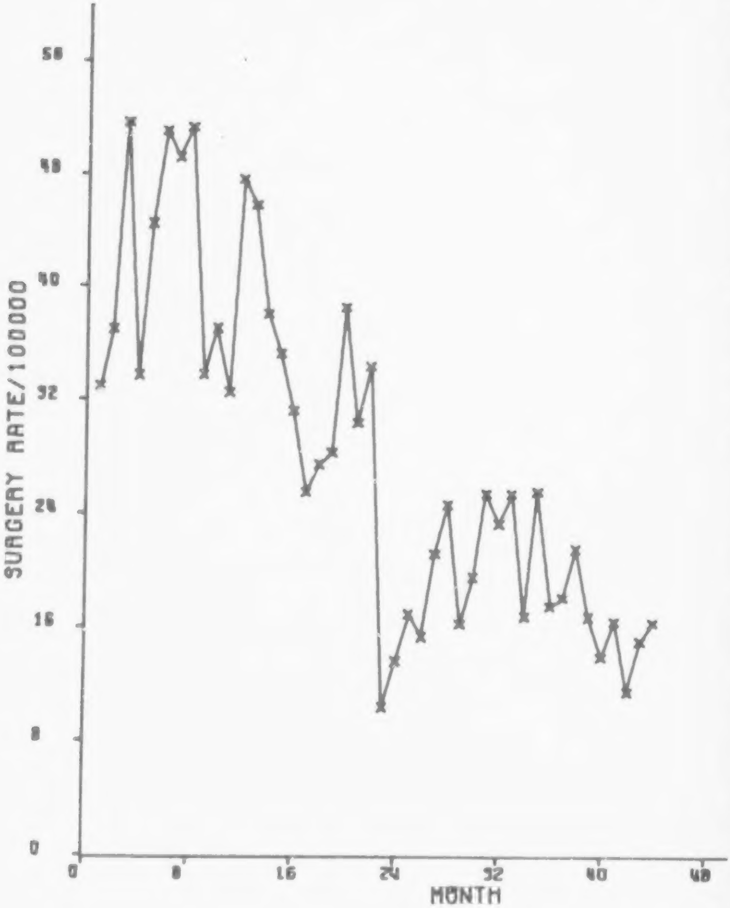


FIG. A-4. Monthly rate of tonsillectomies/adenoidectomies in Bay State.

control area following CPES implementation in Bay State and Central is not statistically significant (p = 0.20 in Bay State and p = 0.15 in Central). There is no indication of a declining trend in hysterectomy rates in the control area (Fig. A-3) or in the pre-CPES period in either region. Thus, these two cases are included in the minimum estimate of the impact of CPES.

Tonsillectomies/Adenoidectomies in Bay State (Fig. A-4)

There are several possible interpretations of this time series. The first is that the fee cut had a delayed impact (from month 9 to month 13), so that the first 13 months of data are eliminated as part of the pre-fee cut period. Surgery rates in the pre-CPES period (months 14–22) appear constant, with a significant drop in rates occurring after CPES (p = 0.00). If the first 13 months are also eliminated from the control, there is no statistically significant drop in rates there (p = 0.42), indicating that the drop in tonsillectomies/adenoidectomies in Bay State is attributable to CPES.

An alternative interpretation is that the rate of tonsillectomies/adenoidectomies

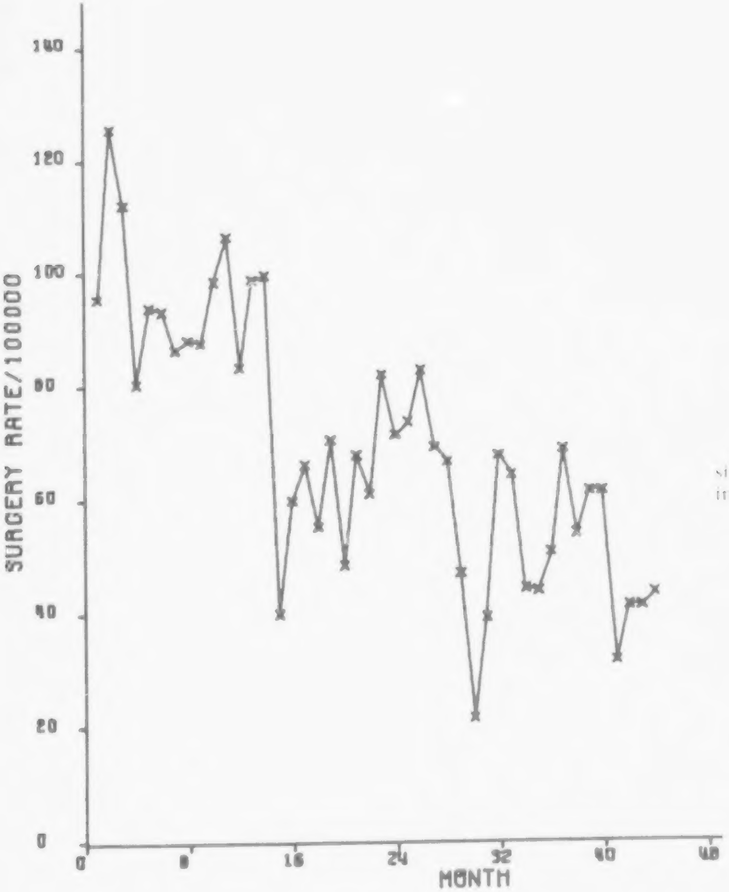


FIG. A-5. Monthly rate of tonsillectomies/adenoidectomies in Central Massachusetts.

was declining regularly over the 44-month period and that, following a temporary drop after CPES implementation, surgery rates returned to the declining trend line. This explanation implies a yearly decline of 21 per cent in surgery rates. In the control area (Fig. A-6), where the base surgery rate was over double that in Bay State, the trend line implies a yearly decline of 11 per cent. Therefore, a 21 per cent decline in Bay State tonsillectomies/adenodectomies is sufficiently unlikely that a declining trend model with no program impact seems unreasonable.

The third possibility is that the model to explain tonsillectomy/adenoidectomy rates should incorporate both a declining trend and a program impact. The yearly decline in rates is 9 per cent, in line with the 11 per cent decline in the control. The program impact predicted by this model (13.62/100,000 less surgeries per month) is almost the same as if the trend is ignored and the pre-fee cut data eliminated (12.52/100,000). Hence, whether one accepts a trend model with an impact or an impact with the trend an artifact of the fee cut and CPES, the impact of CPES is of approximately the same order of magnitude and statistically significant. Thus, this case is included in the minimum estimate.

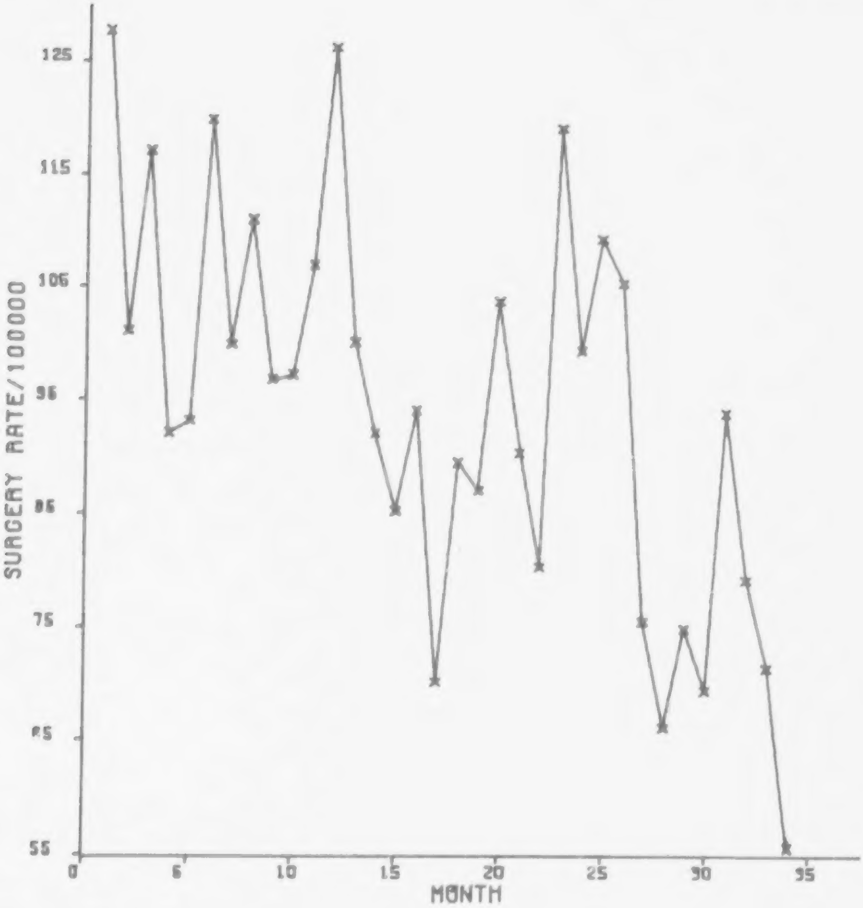


FIG. A-6. Monthly rate of tonsillectomies/adenoidectomies in control area.

Tonsillectomies/Adenoidectomies in Central Massachusetts (Fig. A-5)

If the first 14 months of data are considered part of the pre-fee cut period and eliminated, rates are constant in the pre-CPES period (months 15–29). The post-CPES mean level is lower than the pre-CPES level (p = 0.00). However, even when the pre-fee cut data are eliminated in the control area, the mean in the control area is lower after CPES implementation in Central than before CPES implementation (p = 0.03). A model that postulates a decline over the 44 months of observation implies a yearly decline of 17 per cent, more in line with the 11 per cent decline in the control, particularly given the comparability of base rates. Thus, there is evidence that other factors may have accounted for the drop in Central. Hence, this case is included only in the maximum estimate of program impact.

MEDICAL CARE
January 1982 Vol. XX, No. 1

The Effect of Concurrent Feedback in Reducing Inappropriate Hospital Utilization

JOSEPH D. RESTUCCIA, DR. P.H.

This study, conducted in four general hospitals, evaluates the effectiveness of feedback from utilization review coordinators in reducing the number of inappropriate hospital days. Experimental control was exerted over the type of feedback provided in the utilization review process to produce four treatment groups (including a control group) that vary according to 1) the channel used to inform an attending physician of a probable inappropriately located patient and 2) the amount of discretion permitted the nurse coordinator in deciding whether, when and to whom this information is provided. The results of the study indicate that physician and hospital performance in utilization review, measured both by inappropriate patient days and length of stay, are affected by the feedback strategy employed. For patients inappropriately located for a portion of their hospital stay, provision of concurrent feedback resulted in reduction of approximately two thirds of an inappropriate day and two and one half total days, compared with patients for whom no feedback was provided or for whom physician advisor involvement was required. In addition, the effect of reasons causing patients to be inappropriately located (barriers) was assessed. Barriers outside the realm of influence of the hospital or the physician were found to impede the effectiveness of the utilization review systems.

ONE OF THE EXPLICIT GOALS of the Professional Standards Review Organization (PSRO) program and an implicit goal of other cost-containment programs is to eliminate inappropriate use of health care services.[1] Achievement of this goal assumedly will have beneficial effects on both the cost and quality of health care by eliminating the expense and potential iatrogenic effects of unnecessary services. In addition, resources currently used to provide such services will be made available to fulfill unmet needs. The primary mechanism for the achievement of this goal is utilization review of the necessity for admission and continued hospital stays. Although utilization review systems have been criticized as ineffective and costly,[2-4] there has been little systematic study of the processes that affect the effectiveness of such systems.

The present study was conducted under the premises that 1) conventional review systems are largely ineffective due to their reliance on "bureaucratic" or programmed procedures that limit both the amount of communication between and the discretion of utilization review coordinators and physicians; and 2) the effectiveness of such systems is increased by reliance on "cybernetic" processes that abet communication and discretion. In particular,

---

From the Health Care Management Program, School of Management, Boston University; and the Health Care Research Section, School of Medicine, Boston University.

Supported by National Center, Grant No. 5R18-HSO 1222-02; National Center for Health Services Research, U.S. Department of Health, Education and Welfare; and W. K. Kellogg Foundation Fellowship Award.

Address for reprints: Joseph D. Restuccia, 704 Commonwealth Avenue, Boston University, School of Management, Boston, MA 02215.

0025-7079/82/0100/0046/$01.35 © J. B. Lippincott Co.

this study focuses upon the review coordinator's provision of feedback, the information about occupational performance conveyed to physicians and its relative importance in the effectiveness of hospital utilization review systems.

The study exerted experimental control over the type of information or feedback provided by coordinators to physicians in order to measure the effect of feedback on patients' inappropriate hospital days and length of stay. Before discussing the study, however, it is necessary to clarify the implications of "cybernetic" and "conventional" control mechanisms in the operation of utilization review systems.

Cybernetic Control

According to cybernetic theory, the essential means of achieving control over performance is through the use of feedback. By providing feedback, it is asserted, a system can be made to regulate itself—to exert internal or self-control. As stated by cybernetic theorist Stafford Beer[5]:

> . . . control is not a mandatory exercise, in which people are bullied or things are coerced to operate in a desired way. Rather it is a question of coaxing a system toward optimal performance; or, even better, of arranging for the system to regulate itself . . . (thus) the vital importance of the principle of feedback . . . The job of control . . . (is) less a job for the autocrat as for the steersman.

Feedback, however, is not a naturally occurring phenomenon in social systems. Communication networks must be arranged in particular patterns to insure that feedback will be provided.[6] Figure 1 illustrates how feedback can be provided in a health delivery system. The feedback loop represents information about actual system performance. This information allows comparison of actual performance and desired performance, and consequent adjustment of decisions about service delivery if a discrepancy between desired and actual performance is found. It should be noted that such discrepancies can be caused by events outside the system as well as within the system. The box labeled "uncertain events" represents exogenous variables or events in the system's environment. These events are beyond the control capacity of the system itself, yet may affect system performance in unpredictable ways. However, the feedback loop allows consequent adjustment of the system if such events cause a discrepancy between desired and actual performance. Thus, the system is structured to recognize uncertainty explicitly—to recognize the probabilistic nature of cause and effect.

The concurrent review system evaluated in this study was modeled after the cybernetic system shown in Figure 1. The basic elements of the review process of this system are contained in Figure 2. Figure 3

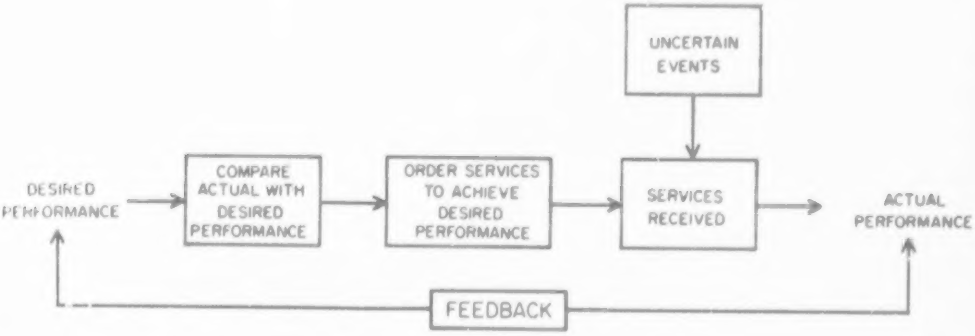

UNCERTAIN EVENTS
DESIRED PERFORMANCE
COMPARE ACTUAL WITH DESIRED PERFORMANCE
ORDER SERVICES TO ACHIEVE DESIRED PERFORMANCE
SERVICES RECEIVED
ACTUAL PERFORMANCE
FEEDBACK

FIG. 1. Cybernetic control system.

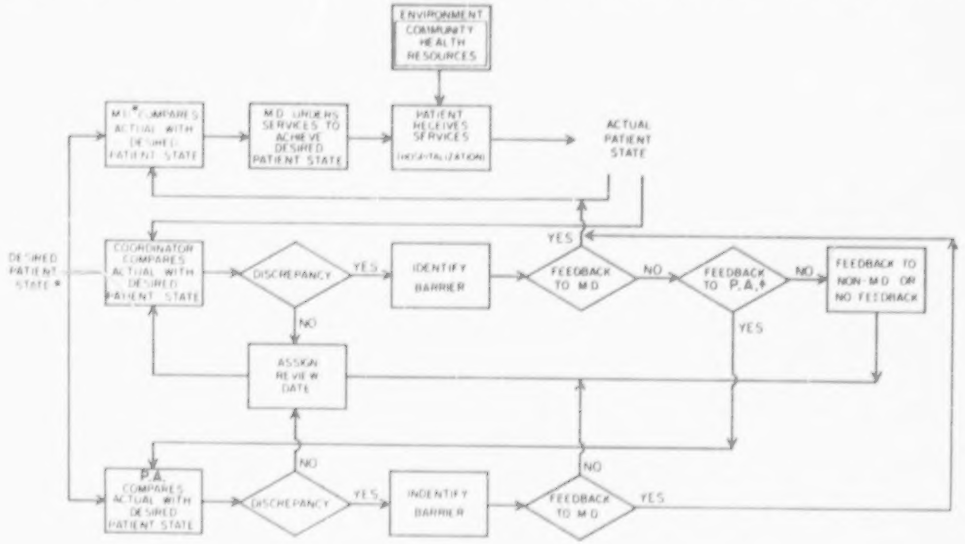
contains the basic elements of a conventional concurrent review system, modeled after federal regulations[7] and the actual procedures used in many utilization review systems.[8] A comparison of these systems reveals that each system represents alternative types of control procedures. The cybernetic system attempts to maintain control by high rates of feedback with discretion, while the conventional system attempts to control performance by programming with santions.

Comparing the "Conventional" and "Cybernetic" Utilization Review Systems

In the following section, each of the three specific hypotheses tested by this study are presented with a discussion of their relevance to the functioning of a cybernetic and a conventional concurrent review system.

Hypothesis 1: The more direct the communication channel from the utilization review coordinator to the attending physician, the more effective is the review system.

The critical element in comparing the cybernetic and conventional review systems is the difference in their communication networks. As shown in Figures 2 and 3, the most obvious difference between the two review processes is that the review coordinator provides feedback directly to the attending physician in the cybernetic system, whereas the conventional system requires that the physician contact a physician advisor if a discrepancy is found (i.e., if the comparison between the desired and actual patient state indicates that hospitalization is not necessary or if necessary hospital services are not received by the patient). If the physician advisor agrees that a discrepancy exists, the attending physician

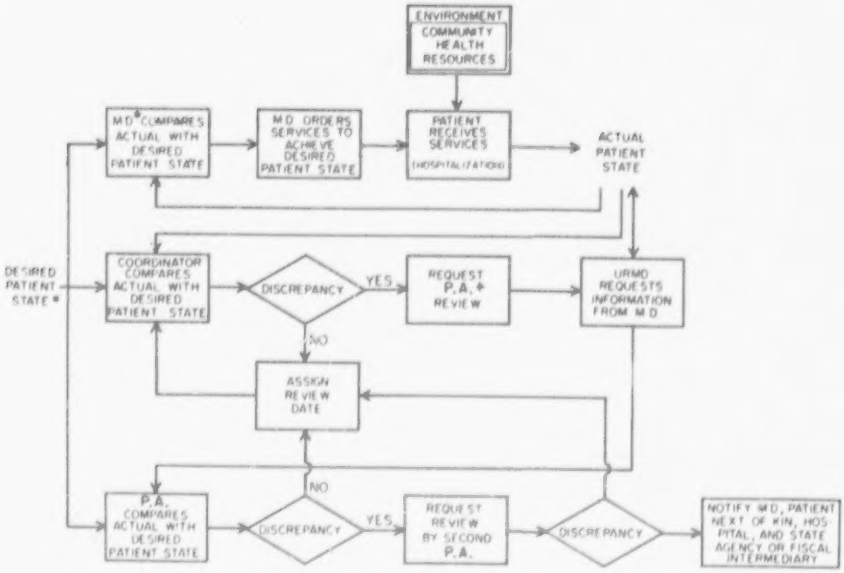

ENVIRONMENT
COMMUNITY HEALTH RESOURCES
M.D.* COMPARES ACTUAL WITH DESIRED PATIENT STATE
M.D. ORDERS SERVICES TO ACHIEVE DESIRED PATIENT STATE
PATIENT RECEIVES SERVICES (HOSPITALIZATION)
ACTUAL PATIENT STATE
DESIRED PATIENT STATE *
COORDINATOR COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
IDENTIFY BARRIER
FEEDBACK TO M.D.
NO
FEEDBACK TO P.A.♦
NO
FEEDBACK TO NON-M.D. OR NO FEEDBACK
YES
YES
NO
ASSIGN REVIEW DATE
NO
NO
P.A. COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
INDENTIFY BARRIER
FEEDBACK TO M.D.
YES

*Appropriate level of care

*"M.D." refers to attending physician

♦"P.A." refers to physician advisor

FIG. 2. Review process in the "cybernetic" utilization review system.


ENVIRONMENT
COMMUNITY HEALTH RESOURCES
M.D.* COMPARES ACTUAL WITH DESIRED PATIENT STATE
M.D. ORDERS SERVICES TO ACHIEVE DESIRED PATIENT STATE
PATIENT RECEIVES SERVICES (HOSPITALIZATION)
ACTUAL PATIENT STATE
DESIRED PATIENT STATE *
COORDINATOR COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
REQUEST P.A.◆ REVIEW
URMD REQUESTS INFORMATION FROM M.D.
NO
ASSIGN REVIEW DATE
NO
P.A. COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
REQUEST REVIEW BY SECOND P.A.
DISCREPANCY
NOTIFY M.D., PATIENT NEXT OF KIN, HOSPITAL, AND STATE AGENCY OR FISCAL INTERMEDIARY

*Appropriate level of care

*'M.D.' refers to attending physician

◆'P.A.' refers to physician advisor

FIG. 3. Review process in the "conventional" utilization review system.

is so informed. Thus, the feedback provided is indirect and attenuated by the likelihood that it is less timely than if provided by the review coordinator directly, due to the requirement that the physician advisor perform a review. In addition, the greater the number of people through which information is transmitted, the greater is the opportunity for the information to be blocked or distorted.[5]

> Hypothesis 2: The more discretion permitted the review coordinator in deciding whether and to whom, feedback will be provided, the more effective is the system.

A second important difference is that the review coordinator in the cybernetic system attempts to identify the reason that a discrepancy exists (i.e., the "barrier" to appropriate utilization of the hospital's beds).[9] He/she determines who or what is responsible for the utilization problem—whether it is the attending physician's pattern of practice, a problem in the hospital's operating systems such as delays in scheduling tests or provision of test results, the patient's lack of cooperation with prescribed treatment or a problem in the hospital's environment such as unavailability of medical services outside the hospital. Therefore, the coordinator must obtain information about the patient's immediate environment, the hospital, and about the hospital's environment, community health care resources. The list of barriers used by the coordinators to identify reasons for inappropriate utilization is contained in Appendix I.

The significance of information about the hospital's environment is that it enables the review coordinator to determine what alternatives to a hospital level of care exist and whether the lack of such alternatives is the cause of the patient remaining in the hospital, despite the discrepancy indicating that the patient could be located elsewhere. For example, the comparison between actual and desired performances may indicate that a patient could be appropriately located in a skilled nursing facility rather than in the hospital. However, the information obtained by the coordinator to determine the cause of this discrepancy may indicate that no skilled nursing facility bed is available, thus making transfer impossible at the time of review. The cybernetic system, then, is structured to recognize uncertainty resulting from environmental events—variables outside the direct influence of the patient care system that nevertheless affect it. In contrast, the conventional system has no explicit provision for considering environmental variables until the review process has reached the stage at which the physician advisor requests further information from the attending physician, "including such factors which preclude treatment of the patient as an outpatient or in an institution providing a lesser level of care."[7]

Similarly, information obtained about the patient's immediate environment may indicate that although a patient may be inappropriately located since he/she is not receiving acute services, the patient may nevertheless require a hospital level of care. For example, the cause of the nonreceipt of services could be failure to provide ordered diagnostic or therapeutic services in a timely manner.

Thus, in the conventional system, the coordinator's decisions are programmed or restricted by a deterministic plan. He/she is little more than a monitor of performance, responsible merely for determining if a discrepancy exists and for referring discrepancies to a physician advisor for further review and resolution. In the cybernetic system, the coordinator assists in the resolution of the problem. By identifying the cause of the discrepancy, the coordinator may begin to develop a stategy to resolve it. The information about the discrepancy and this strategy can then be fed back to the attending physician. Similarly, it can be fed back to others who might assist in resolving the problem, such as discharge planners, social workers, administrators, medical consultants, nurses or personnel from relevant diagnostic or therapeutic departments. In the cybernetic system the coordinator becomes a problem solver as well as a monitor. He/she is accorded sufficient discretion to decide among multiple alternative actions: whether and when to contact the attending physician, physician advisors, other hospital departments or personnel, or any combination he/she chooses.

Hypothesis 3: An effective system is achieved through provision of feedback by coordinators to attending physicians without the use of the sanction that is implied by the involvement of a physician advisor.

Another important difference between the cybernetic and conventional systems is that the former relies on transmission of information and the discretion of attending physicians in acting on this information, whereas the latter primarily employs negative sanctions to maintain control. In the cybernetic system, control occurs as a result of information indicating a discrepancy in system performance. The coordinator informs the attending physician that the patient *may* be "inappropriately" hospitalized, i.e., does not appear to require hospitalization or to be receiving services that would warrant continued hospitalization. Based on this information, and on the problem-solving assistance initiated by the coordinator, the attending physician presumably will take one of the following actions: 1) ensure that the patient's location be made consistent with the level of care required (i.e., discharge or transfer of the patient); 2) order the services required for the patient to be appropriately located in a hospital (i.e., order acute services available only in a hospital setting); or 3) provide further documentation justifying the appropriateness of hospitalization. Typically, the coordinator will request intervention of a physician advisor only if the attending physician fails to respond adequately to the information provided. Thus, appeal to a physician advisor is invoked only as a backup or failsafe control mechanism.

In contrast, the conventional system requires intervention of a physician advisor whenever a discrepancy is found, an intervention that is likely to be viewed as a sanction. The coordinator is invoking a "higher" authority; information to the attending physician is provided by someone of comparable occupational status—a peer rather than a subordinate. Moreover, this information is not necessarily coupled with an offer of assistance to resolve whatever problem exists. The conventional review system operates under the assumption that the attending physician alone has the responsibility and capability of resolving the problem, yet it does not offer him/her the opportunity to achieve resolution without prior evaluation and intervention by another physician. This system assumes that the physician advisor, a source of external control, must influence the attending physician to resolve the problem and that the physician advisor will be motivated to exert this influence. It fails to consider explicitly the responsibility of others for the existence and resolution of the problem or to consider the potential effect of environmental barriers in causing the problem. Little or no attempt is made to coordinate hospital and community resources that might assist attending physicians to locate patients appropriately. Thus, the conventional review system primarily relies upon programmed decisions with sanctions to achieve control, while the cybernetic system employs feedback with discretion.[10, 11]

Methods

The study was conducted in four general, acute-care hospitals in the San Francisco Bay area. Table 1 provides summary characteristics of each hospital. For a two-month period in each of the four hospitals, nurse review coordinators were asked to provide different types of feedback to physicians about patients randomly assigned to one of four treatment groups. During this period, physicians, except the utilization review committee chairmen and directors of medical education, were not informed of the study to minimize reactive effects that might threaten external validity. Study patients were limited to those covered by Medicare and admitted to a medical or surgical service at one of the four hospitals. The assignment of patients was performed by the coordinators' secretaries prior to the admission review according to a systematic random sampling design.[12]

The four treatment groups represent feedback strategies that vary according to the channel used to provide feedback to physicians and the amount of discretion accorded to the coordinator in deciding whether, when and to whom feedback was provided. In each strategy the feedback decision was triggered by the coordinator's detection of a possible inappropriate patient day according to the application of the Medicare level-of-care criteria, presented in Figure 4.[13] Methodological tests of these criteria indicated that application by registered nurses trained in their use resulted in reliable (i.e., consistent) and valid judgments about appropriateness of hospital utilization.[14] Each of the six coordinators at the four study hospitals was a registered nurse who had training and experience in applying these criteria.

The four feedback strategies are as follows:

1. Direct Feedback (DF): The coordinator directly informs the attending physician that his/her patient appears to be inappropriately located.

2. Indirect Feedback (IF): The coordinator informs a physician advisor that a patient appears to be inappropriately located. It then becomes the physician advisor's responsibility to decide whether to contact the patient's attending physician.

TABLE 1. Summary Description of Study Hospitals

Hospital	Sponsorship	No. of Beds	Teaching Status	1975 Occupancy Rate (%)	Occupancy Rate During Study (%)	% of Medicare Patients	% of Medicaid Patients
1	Community	214	None	79.4	80.1	40	34
2	Religious	254	None	70.1	74.1	41	23
3	Community	437	Affiliated	78.2	74.3	36	21
4	County	584	Affiliated	57.0	56.7	21	54

3. No Feedback (NF): Neither the attending physician nor a physician advisor is informed that a patient appears to be inappropriately located.

4. Judgmental Feedback (JF): The coordinator is permitted the discretion to decide whether and in what order to inform the attending physician, physician advisor, both, or neither physician. Thus, this category may include any one or any combination of the three categories above, depending upon the judgment of the coordinator.

The four strategies were structured to represent different methods of control. Indirect feedback is closest to programming with sanctions, since it permits the coordinator no discretion and requires involvement of a physician advisor. The coordinator must follow a plan to the extent that he/she must involve a physician advisor and a sanction is invoked by the physician advisor's intervention. Judgmental feedback, in contrast, is closest to high feedback with discretion, since the coordinator's feedback decision is not programmed but discretionary, and since sanctions over the attending physician's decision are rarely invoked (assuming that the coordinator will typically choose to provide either direct or no feedback prior to or instead of providing indirect feedback).

Direct feedback is representative of an intermediary position. No sanctions are invoked, since the attending physician is informed of a patient's inappropriate location without involvement of a peer, but neither is the coordinator permitted discretion in the feedback decision. The category in which no feedback is provided serves as a control group, since no concurrent action is taken by the coordinator for inappropriately located patients in this category.

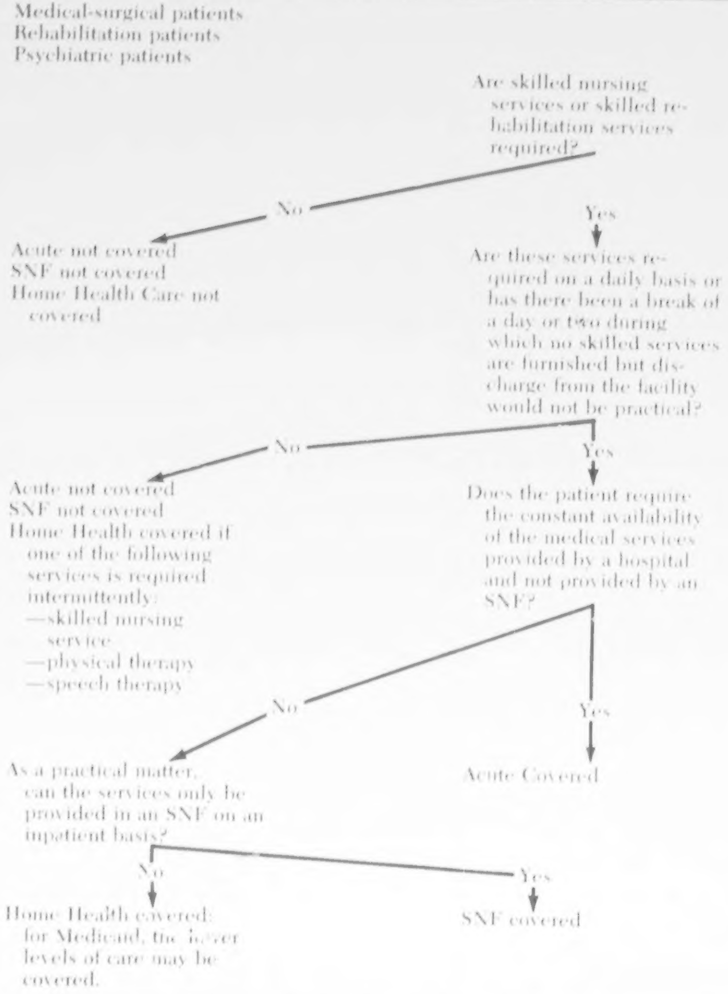
Dependent Variables

The primary dependent variable is the number of inappropriate patient days occurring for patients who experience one or more inappropriate days during their hospital stay. The variable is a direct measure of the degree of control over physician performance exerted by each "treatment" or feedback strategy in complying with the goal of eliminating inappropriate hospital utilization. The null hypothesis, then, is that there is no difference in the average number of inappropriate days in each treatment category.

A second dependent variable is the total case mix adjusted length of stay for patients in each treatment group. While a less sensitive measure of control exerted by the utilization review system than number of inappropriate days, a reduction in inappropriate days would likely result in a reduction in patients' total hospital stay.

Standardizing Patient Reviews

An important variable in utilization review is prediction of the day of the patient's stay on which review should be conducted. Studies have demonstrated that the proba-


Medical-surgical patients
Rehabilitation patients
Psychiatric patients
Are skilled nursing services or skilled rehabilitation services required?
No
Yes
Acute not covered
SNF not covered
Home Health Care not covered
Are these services required on a daily basis or has there been a break of a day or two during which no skilled services are furnished but discharge from the facility would not be practical?
No
Yes
Acute not covered
SNF not covered
Home Health covered if one of the following services is required intermittently:
—skilled nursing service
—physical therapy
—speech therapy
Does the patient require the constant availability of the medical services provided by a hospital and not provided by an SNF?
No
Yes
As a practical matter, can the services only be provided in an SNF on an inpatient basis?
Acute Covered
No
Yes
Home Health covered: for Medicaid, the lower levels of care may be covered.
SNF covered

Examples

Skilled Nursing and Skilled Rehab Services

1. Overall Management and Evaluation of Care Plan.
2. Observation and Assessment of the Patient's Changing condition.
3. Patient Education Services.
4. Direct Nursing Services
   a. IV, IM, or SC injections and IV feeding;
   b. Levin tube and gastrostomy feeding;
   c. Nasopharyngeal and tracheotomy aspiration;
   d. Insertion and new sterile irrigation and replacement of catheters;
   e. Application of dressings involving prescription medications and aseptic techniques;
   f. Treatment of extensive decubitus ulcers or other widespread skin disorder;
   g. Heat treatments;
   h. Initial phases of a regimen involving administration of medical gases;
   i. Rehabilitation nursing procedures, including the related teaching, and adaptive aspects of nursing, that are part of active treatment.
5. Direct Rehabilitation Services
   a. Therapeutic exercises or activities;
   b. Gait evaluation and training;
   c. Range of motion exercises;
   d. Maintenance therapy;
   e. Ultrasound, short-wave and microwave therapy treatments;
   f. Hot pack, hydrocollator, infrared treatment paraffin baths and whirlpool;
   g. Speech pathologist or audiologist services.

From the Federal Register, Vol. 40 (186), September 24, 1975.

FIG. 4. Medicare levels-of-care criteria.

bility of a patient day being inappropriate generally increases as length of stay increases.[15] Thus, it is necessary to "standardize" review dates to eliminate any bias (unconscious or otherwise) between treatment groups that might be introduced by the coordinators' assignment of review dates.

Standardization of review dates was achieved by employing strict length-of-stay-by-diagnosis guidelines to dictate when reviews should be made. Reviews were made in accordance with provisions of federal utilization review regulations applicable at the time of the study.[7] In general these requirements are that review be performed with 24 hours of admission and on specified percentiles (50th, 75th, 90th) of length of stay by diagnosis for a region, with a maximum of five working days between reviews irrespective of length-of-stay-by-diagnosis guidelines.

In addition, if the coordinator decided that a patient was inappropriately located, the patient was reviewed on each subsequent day of stay until his/her discharge. On each day that a review was made, the patient's level of care was determined according to the Medicare levels-of-care criteria. These consecutive reviews were necessary to measure accurately the dependent variable, the number of inappropriate days per patient. It could have been assumed that once an inappropriate day was detected, the patient would remain inappropriately located on each subsequent day of stay until discharge. However, if the patient became appropriately located in the hospital on one of these subsequent days, the dependent variable would have been erroneously inflated.*

Limitations on the Experimental Design

Two limitations were placed on the experimental design to minimize risking loss of Medicare revenue due to an increase in appropriate use as a result of the study.† First, the random assignment of patients to treatment categories was conducted so as to yield twice the proportion of patients in the Judgmental Feedback group as in any one of the other three treatment groups. Second, a limit was imposed on the length of time an inappropriately located patient was allowed to remain in the three treatment categories other than the Judgmental Feedback group. This time limit was four consecutive inappropriate days, including the first inappropriate day detected by the coordinator. Thus, if a patient in one of these three treatment categories remained inappropriate (i.e., was not discharged or had not experienced a rise in level of care that justified continued hospitalization) for four consecutive days, the coordinator was permitted to ignore the treatment category assigned and take whatever action he/she deemed necessary to prevent further inappropriate utilization. In effect, the coordinator could switch to the judgmental strategy on the fourth consecutive inappropriate day.

The first limit, the unequal sizes of the treatment groups, causes the experimental design to be nonorthogonal, or unbalanced. However, this does not cause any serious problems in analyzing results since it is possible to adjust statistically for nonorthogonal treatment effects.[16] The second limit is potentially more serious. if the mean response time of the treatment effects as measured by the dependent variable (i.e., the number of days between detection of an inappropriate day and resolution of the utilization problem) were to exceed four days, no differences would be detected among the four treatment categories; the alternative hypothesis (i.e., that significant differences exist among treatments) might be true, but might require more than four days for differences to occur. At a minimum, this four-day limit could be expected to affect results by decreasing the differences in response time that would normally occur between feedback strategies without the four-day limit. Thus, this limitation subjected the alternative hypothesis to a severe test by, in effect, decreasing the power of any statistical test to detect significant differences among feedback strategies.

---

* This "rise" from a nonacute level of care was a distinct possibility when the barrier was one related to inefficient delivery of service and the patient was assigned to a noncontrol treatment category. For example, if the barrier were a delay in a diagnostic test, the coordinator's feedback to a physician might incite corrective action (e.g., facilitate receipt of the test results) and the patient's level of care could again become appropriate. If the coordinator had not continued to review the patient, such appropriate day(s) subsequent to an inappropriate day(s) would have gone undetected. Thus, when such a rise in level of care was detected, the coordinator resumed reviews in the same manner as for a newly admitted patient.

† At the time this study was conducted, if the rate of retroactive denial of claims exceeded 5 per cent of the total Medicare patient days in any three-month period, the hospital lost its Medicare "waiver of liability" and payment for denied claims was withheld.

In addition to the limitations purposely imposed on the experimental design, other factors tended to reduce the differences among the treatment groups. These differences are likely to be reduced to the extent that anticipatory socialization of physicians occurs as a result of their awareness that the hospital has a functioning concurrent review system. Thus, their behavior is, in part, conditioned by the past history of the hospital review system regardless of feedback strategy. In addition, some control might be exerted as a result of potential, retroactive denial of payment for patients that the Medicare fiscal intermediary judged inappropriately located on the basis of claims and records review. This sanction, however, usually occurs at least two months after the claim is submitted and is an infrequent event (rarely having exceeded 2 per cent of total Medicare patient days in the study hospitals).

To the extent that differences between treatment categories are reduced by anticipatory socialization, retroactive denials and limitations imposed on the experimental design, the study hypotheses are subjected to a more difficult test. If differences in the number of inappropriate days nevertheless occur, there is strong evidence that they are attributable to the differences in the feedback strategies.

Results

In the four participating hospitals, the coordinators reviewed a total of 1,456 Medicare medical and surgical service patients during the study. Three hundred twenty-four patients (22.3 per cent) were found to have had at least one inappropriate day, with an average of 2.85 inappropriate days (21.1 per cent), in their average total length of stay of 13.49 days.As would be expected, the average length of stay of patients with inappropriate days was considerably longer than that of patients with no inappropriate days: 13.49 days (4,371 days/324 patients), compared with 9.31 days (10,540 days/1,132 patients). The average length of stay of all patients reviewed was 10.24 days, a figure approximately equal to the average length of stay of all Medicare patients in the four study hospitals during the year of the study.

Effect of Feedback Strategies

To determine the effect of feedback strategy on the dependent variable, the number of inappropriate patient days (IDAY), an analysis of variance (ANOVA) was conducted on 324 of the 333 patients who experienced at least one inappropriate day (nine patients who died during the study were excluded from the analysis). In order to focus on the effect of the feedback strategy provided, two independent variables were included in the ANOVA: the feedback strategy and the hospital. Inclusion of hospitals (with each hospital constituting one level) as a second independent variable removed variance in the dependent variable attributable to differences between the hospitals. In addition, a log transformation was performed on the dependent variable so that it more closely conformed to a normal distribution. Results are reported for both the transformed (IDAY) and untransformed (LGIDAY) dependent variable: the former for statistical inference, the latter for estimation of treatment effects.[16]

As shown in Table 2, the difference in LGIDAY among the four feedback strategies is statistically significant. IDAY in the judgmental and direct feedback strategies is approximately one half day less than in the control group. The indirect strategy, in contrast, resulted in no reduction in inappropriate days compared with the control group.[‡]

A two-way ANOVA, with feedback strategy and hospital as independent variables, was also performed to determine the effect of feedback on average length of stay (LOS). To adjust for case mix differences among patients, the Western United States PAS region average length of stay according to final diagnosis, age group, presence or absence of surgery, and presence or absence of multiple diagnoses was used as a covariate.[17]

In addition, a log transformation was performed on length of stay to approximate a normal distribution of this dependent variable (LGLOS). The results of this analysis parallel those that used inappropriate days as a dependent variable, with the exception that the direct feedback strategy was less influential in reducing total hospital stay than in reducing inappropriate days. As shown in Table 2, the major difference in length of stay is between the judgmental feedback strategy and both the indirect strategy and the control group.[§]

Effect of Feedback Strategies Excluding Patients with Environmental Barriers

To determine the extent to which environmental barriers impeded the effectiveness of feedback strategies, an ANOVA on LGIDAY and LGLOS was conducted excluding those patients who had inappropriate days associated with barriers beyond the immediate control of the physician, hospital, patient and patient's family. As shown in Table 3, the mean number of inappropriate days for the 192 patients without environmental barriers is 2.70, compared with 2.85 for all 324 patients with inappropriate days (and 3.07 for the 132 patients with environmental barriers).

The exclusion of patients with environmental barriers results in the same pattern of means found with all patients included, but with larger differences and higher significance levels between means. IDAY in the judgmental and direct strategies, respectively, is approximately 1.5 days and one day less than in the control and indirect feedback groups, while no difference was found between the control group and the indirect strategy.[‖] The difference of over 3.5 days in length of stay between the judgmental and control groups shows the influence of the coordinator's discretion in reducing length of stay, particularly when not contrained by environmental barriers. The provision of direct feedback also exhibits an average length of stay in the hypothesized direction, although differences with other group means are not statistically significant.[#]

Comparison of Hospitals

Because of the small sample size of hospitals ($N = 4$) and the impossibility of controlling interorganizational characteristics

---

[‡] Scheffe's test applied to the contrasts JF–NF, JF–IF, DF–NF and DF–IF are each significant at the $p \leq 0.05$ level. In other words, the differences are primarily attributable to the lower number of inappropriate days in the judgmental and direct categories as compared with the indirect category and control group.

[§] Scheffe's test applied to the contrast JF – (IF + DF)/2 is significant at the $p \leq 0.10$ level.

[‖] Scheffe's test indicates differences in inappropriate days excluding environmental barriers at the $p \leq 0.01$ level for contrasts JF–NF, JF–IF and DF–IF and at the $p \leq 0.05$ level for the contrast DF–NF.

[#] Scheffe's test indicates length-of-stay differences at the $p \leq 0.05$ and $p \leq 0.01$ levels for the contrasts JF–IF and JF–NF, respectively.

TABLE 2. Average Number of Inappropriate Days and Length of Stay by Feedback Strategy

Strategy	N	IDAY*	LGIDAY*	LOS†	LGLOS†
Judgmental	145	2.59	0.48	12.23	1.05
Direct	68	2.75	0.48	13.81	1.10
Indirect	60	3.25	0.56	15.23	1.14
None (control)	51	3.25	0.57	14.59	1.12
Total	324	2.85	0.51	13.49	1.09
			F = 2.88 P ≤ 0.04		F = 3.10 P ≤ 0.03

* Adjusted for hospital differences.

† Adjusted for hospital and covariate differences.

that might affect inappropriate utilization in the study hospitals, conclusions drawn from a comparison of hospital differences must be regarded as tentative. However, certain consistencies with the results of the comparison by feedback strategy allow inferences to be made about the causes of the hospital differences.

Table 4 presents the average number of inappropriate days and the average length of stay by hospital. The ANOVA on the log of the values in Table 4 showed no significant differences among hospitals in the number of inappropriate days, whether for all patients (F = 1.88; N.S.) or for those without environmental barriers (F = 1.11; N.S.). Length of stay adjusted for case mix (again using regional average length of stay by diagnosis as a covariate) was found to be different for all patients (F = 11.71; p ≤ 0.000) and for those without environmental barriers (F = 6.93; p ≤ 0.000). The length-of-stay differences are primarily due to the shorter stay at Hospital 1. Also, despite no significant differences in inappropriate days among hospitals, Hospital 1 had fewer inappropriate days than did the other three hospitals.

The relative effectiveness demonstrated by Hospital 1 might be attributable to such factors as occupancy rate pressures, types of patients or modes of medical staff practice. Those factors are particularly likely to be relevant in comparing Hospital 1 with Hospital 4. Hospital 1 is a medium-sized, community general hospital with an occupancy rate of 80 per cent. Hospital 4 is a large, county general hospital with an occupancy rate of 57 per cent. Hospital 4, then, would have less pressure to dis-

TABLE 3. Average Number of Inappropriate Days and Length of Stay by Feedback Strategy Excluding Patients with Environmental Barriers

Strategy	N	IDAY*	LGIDAY*	LOS†	LGLOS†
Judgmental	84	2.07	0.43	11.13	1.02
Direct	43	2.58	0.46	13.18	1.10
Indirect	40	3.64	0.60	13.68	1.11
None (control)	25	3.51	0.57	14.89	1.13
Total	192	2.70	0.49	12.61	1.07
			F = 5.62 P ≤ 0.001		F = 3.41 P ≤ 0.02

* Adjusted for hospital differences.

† Adjusted for hospital and covariate differences.

TABLE 4. Average Number of Inappropriate Days and Length of Stay by Hospital

Hospital	All Patients with Inappropriate Days Included			Patients with Environmental Barriers Excluded		
	N	IDAY*	LOS†	N	IDAY*	LOS†
1	54	2.26	12.06	47	2.17	9.01
2	106	2.99	13.32	49	2.91	13.69
3	121	2.72	14.26	70	2.91	13.69
4	43	3.61	15.21	26	2.70	14.18
Total	324	2.85	13.49	192	2.70	12.61

* Adjusted for hospital differences.

† Adjusted for hospital and covariate differences.

charge patients quickly or resolve problems of inappropriate utilization due to a need for beds. In addition, being a county hospital, it has a patient mix with a high proportion of medical, social and/or financial problems that would make appropriately locating such patients difficult. Finally, the extensive teaching programs of Hospital 4, as well as other differences in the composition of its medical staff, may result in a less effective utilization review system than that of Hospital 1.

Without further investigation we can only speculate about whether and to what extent these factors contribute to inappropriate utilization. Moreover, the difference between Hospital 1 and Hospitals 2 and 3 are not as pronounced relative to such factors as the differences between Hospital 1 and Hospital 4. However, based on the results of the comparison of strategies, there are factors that can be examined to explain (at least in part) differences in the relative effectiveness of the utilization review systems in the four hospitals.

First, our results show that direct feedback is more effective than indirect or no feedback. Thus, if there are differences between the hospitals in the amount of direct feedback provided to the physician, we would expect this to be reflected in the measures of utilization review system performance. Second, the more discretion permitted (or assumed by) the coordinator, the more likely is the utilization review system effectiveness. Finally, the greater the proportion of environmental barriers causing inappropriate utilization, the less control the utilization review system is capable of exerting and, thus, the more likely is inappropriate tuilization.

Table 5 presents measurements of these three factors. Column 1 indicates the feedback strategy chosen by the coordinators at each hospital for patients assigned to the Judgmental Feedback group. At Hospital 1, the proportion of direct feedback provided was the highest of the four hospitals, while the proportion of times in which no feedback was provided was the lowest.

Column 2 indicates the percentage of cases in which the coordinator did not adhere to the study design, primarily by failing to provide feedback before expiration of the four-day time limit. Nonadherence was substantially higher in Hospitals 2 and 3 than in Hospitals 1 and 4. The major reason for this difference is that in Hospitals 2 and 3, the coordinators' work load exceeded their capacity to meet it. This reason also helps to account for the differences in feedback strategy chosen by the coordinators (column 1). At Hospitals 2 and 3, the coordinators' excess work load precluded their exercising discretion in providing feedback as often and as promptly as they would otherwise have chosen to do so. Further evidence of this problem is that shortly after the study, both Hospitals 2 and 3 added additional coordinators to handle the work load.

Finally, as shown in column 3 of Table 5, the proportion of environmental barriers is considerably lower in Hospital 1 than in the other three hospitals. Moreover, the proportion of physician barriers is considerably higher—precisely the type of barriers we would expect to be most amenable to resolution by providing feedback to physicians. In part, this low proportion of environmental barriers is due to more available long-term care facilities and home health services in the geographic service area of Hospital 1 than in the areas of the other study hospitals. In addition, as indicated in column 4 of Table 5, Hospital 1 had more experience providing feedback to attending physicians and reducing barriers to appropriate utilization through, for example, increased efforts in discharge planning.[9] Thus, it is not surprising that its utilization review system manifests a high degree of effectiveness.

It remains to be explained, however, why the utilization review system of Hospital 4 is not more effective. Of the four study hospitals, Hospital 4 has the highest average number of inappropriate days and the longest length of stay (Table 4). Yet, on the factors predicting utilization review system effectiveness (Table 5), it ranks second only to Hospital 1.

The reason for this apparent discrepancy was touched upon earlier when comparing Hospitals 1 and 4. As a county general hospital, Hospital 4 tends to receive the most difficult utilization problems—patients with little or no financial resources to pay for medical care and social problems that make appropriate location difficult. These problems should be reflected in the proportion of environmental barriers encountered. Yet this proportion is actually lower than and equal to those of Hospitals 2 and 3, respectively. However, when we look at multiple barriers, we find that 23 per cent

TABLE 5. Predictors of Study Hospitals' Utilization Review System Performance

Hospital	Type of Feedback* Provided for Patients in Judgmental Group (% of Patients)		Adherence to Study Design (% of Patients)		Distribution of Barriers by Area of Responsibility (% of Patients)		No. of Years Using "Cybernetic" UR System
1	Direct	65	Adherence	91	Physician	61	3
	Indirect	12	Nonadherence	09	Hospital	08	
	None	23			Patient	13	
					Environment	17	
2	Direct	37	Adherence	76	Physician	38	0.50
	Indirect	4	Nonadherence	24	Hospital	03	
	None	59			Patient	03	
					Environment	57	
3	Direct	33	Adherence	71	Physician	42	0.25
	Indirect	2	Nonadherence	29	Hospital	03	
	None	65			Patient	08	
					Environment	47	
4	Direct	50	Adherence	91	Physician	41	2
	Indirect	20	Nonadherence	09	Hospital	06	
	None	30			Patient	06	
					Environment	47	

* In cases in which the coordinator provided both direct and indirect feedback, only the type provided first was included.

of the inappropriately located patients at Hospital 4 had more than one barrier that required resolution before the patient could be appropriately located. In Hospital 1, 2 and 3, this proportion was only 5, 6 and 7 per cent, respectively.

In addition, the environmental barriers encountered by Hospital 4 were less amenable to resolution than were those at the other hospitals. This is indicated by the difference in inappropriate days found when including and excluding patients with environmental barriers: 0.91 days for Hospital 4 (3.61 − 2.70) compared with −0.09, 0.08, and −0.19 in Hospitals 1, 2 and 3, respectively (Table 4). Thus, environmental penetration—factors beyond the control of the hospital or its utilization review system—account for much of the relative ineffectiveness exhibited by the utilization review system at Hospital 4.

Summary and Discussion

For all 1,456 patients reviewed in this study, the average difference in length of stay between the Judgmental Strategy and the control group is slightly more than one half day per patient (22.3 per cent of 1,456 patients with inappropriate days × 14.59 days − 12.23 days = 0.52 days/patient). The average direct cost of the utilization review systems at the four hospitals was less than $10 per discharge. If the marginal cost savings associated with a reduction of one half day per patient in length of stay exceeds $10, the utilization review systems at these hospitals have a positive benefit–cost ratio. Using an HEW estimate of marginal cost savings per Medicare day of $87.00 for the San Francisco PSRO area in 1977,[18] the benefit–cost ratio is in excess of 4 ($87.00/2: $10.00 > 4 : 1). Furthermore, in this study no attempt was made to "focus" review on patients with a high probability of experiencing inappropriate days. Similarly, no attempt was made to select review dates within a patient's length of stay (other than percentile of length of stay by diagnosis) to maximize the probability of detecting an inappropriate day. Lower review costs can be achieved with such systems for focused review[8] and review date selection.[14]

In general, the study hypotheses, although slightly revised, have been confirmed.

Hypothesis 1: The more direct the feedback channel, the greater is the control over professional performance.

When coordinators directly informed attending physicians that a patient appeared to be inappropriately hospitalized, the physicians acted to resolve utilization problems more rapidly than when no information was given or when coordinators involved a physician advisor. As previously discussed, direct feedback conveys more timely and accurate information. In addition, as Blau and Scott[19] assert, high rates of communication tend to improve decision-making processes through 1) relief of anxiety about decisions resulting from social support; 2) provision of error correction resulting from different approaches of different people to solving a problem; and 3) incentives for advancing good suggestions and criticizing poor suggestions resulting from competition for respect in judgments. The consequent increase in ability to solve problems is particularly relevant in this situation since explicit in the coordinator's feedback is his/her assistance in identifying and resolving the utilization problem.

Hypothesis 2: The more discretion allowed in the feedback decision, the greater is the control over professional performance.

When coordinators were permitted to decide whether, when and to whom feedback was provided, the time to resolve utilization problems was shorter than when they were restricted to providing a particular type of feedback. A similar finding was made by Gertman and Egdahl[8] in a case study of utilization review activities in 44 Massachusetts hospitals.

One reason for this finding is that the coordinators preferred the responsibility and enhanced professional status associated with discretion and were thus motivated in their job performance. A second reason is that the coordinators felt confidence in the levels-of-care criteria and consequently in their appropriateness decisions made by applying these criteria. In effect, the criteria helped the coordinators to decide when discretion was warranted and helped to increase their confidence in exercising discretion. Furthermore, it is likely that limiting the coordinators' repertoire of potential responses to utilization problems frequently resulted in inappropriate responses. On one extreme, for example, the coordinator might encounter a routine case of inappropriate utilization, such as a patient kept for an additional day waiting for transfer to a skilled nursing facility. To require any response at all (other than reviewing the patient the next day to see if the transfer has occurred) would be unnecessary and, possibly, an impetus for physician resistance. On the other hand, the coordinator may encounter a case with a serious barrier to appropriate utilization, such as a delay in performing a critical test or procedure. In this situation, the attending physician—and possibly others involved in the patient's treatment—should be informed, the cause of the delay identified and corrective action taken as soon as possible.

Hypothesis 3: An effective system is achieved through provision of feedback by coordinators to attending physicians without the use of the sanction that is implied by the involvement of a physician advisor.

The relative effectiveness of direct and judgmental feedback supports this hypothesis. The relative ineffectiveness of indirect feedback indicates that involving a physician advisor caused no control to be exerted over attending physicians' performance. These results call into question the conventional concept of "peer review" which asserts that only a peer (e.g., a physician) can review and positively effect change in the behavior of other peers. This is not intended to imply that peer review is unnecessary, but rather that in level-of-care determinations a nonphysician, guided by explicit criteria and medical staff support, can positively influence physician performance. In addition, the degree of influence is greater than that exerted by a conventional or "bureaucratic" system that requires peer intervention. Thus, *routinely* involving physician advisors to corroborate the coordinator's level-of-care decision and to provide feedback to attending physicians is not only unnecessary but detrimental to utilization review system performance. Physician advisors at all four study hospitals were reluctant to confront their peers when informed of utilization problems, a finding similar to that of Sanazaro and Worth[20] in a study of concurrent quality assurance in five PSRO areas. A more effective and certainly less costly use of physician advisor time would be for the physician advisor to act as a consultant employed to advise and assist the coordinator in cases seen by the coordinator as requiring the expertise or authority of a physician.

The study, then, has demonstrated that self-regulation of professional performance can be achieved. Requisite to self-regulation is rendering visibility to performance. This visibility, in turn, requires an information system capable of comparing desired and actual performance and of providing timely feedback about discrepancies in performance. When such information was provided—along with sufficient discretion to the provider of the information, identification of the causes of discrepancies (i.e., barriers) and problem-solving assistance—physicians typically responded with action that corrected the discrepancy, without resort to sanctions. This finding is particularly relevant to hospitals, organizations having little access to sanctions over physicians, the professional group upon whom they are most dependent yet whose performance they are expected to control.

The potential amount of control that can be exerted, however, is also dependent upon the extent to which barriers are within the scope of physicians' and hospitals' realms of influence. Environmental penetration—intrusion of social, economic and financial barriers into the confines of the hospital—restricts the ability of control systems to achieve desired performance. A significant amount of control, then, can be exerted by a system operating on the basis of cybernetic principles, with the upper limit on the amount of control determined by the extent of environmental penetration.

Acknowledgments

The author is indebted to the utilization review coordinators and the medical and administrative staffs of the four study hospitals, and to Don C. Holloway, Ph.D., for his advice and encouragement.

References

1. Goran MH. The evolution of the PSRO Hospital Review System. Med Care 1978;17(Suppl):1.

2. Averill MA, McMahon LF. A cost benefit analysis of continued stay certification. Med Care 1977;16:158.

3. Dobson A, et al. PSROs: their current status & their impact to date. Inquiry 1978;15:113.

4. McSherry CK. Quality assurance: the cost of utilization review and the educational value of medical audit in a university hospital. Surgery 1976;80:122.

5. Beer S. Decision and control. London: Wiley, 1966:255.

6. Warner DM, Holloway DC. Frameworks for quantitative decision making & cybernetic control. In: Decision making and control for health administration. Ann Arbor: Health Administration Press, 1978.

7. U. S. Department of Health, Education and Welfare, Social and Rehabilitation Service. Utilization review. Federal Register March 30, 1976;41:13452.

8. Gertman PM, Egdahl RE. The dynamics of utilization review: a case study of 44 Massachusetts hospitals. Ann Surg 1978;188:544.

9. Restuccia JD, Holloway DC. Barriers to appropriate utilization of an acute facility. Med Care 1976;14:559.

10. March JG, Simon HA. Organizations. New York: Wiley, 1958.

11. Hage J. Communication and organizational control: cybernetics in health and welfare settings New York: Wiley, 1974.

12. Hess I, Reidel DC, Fitzpatrick TB. Probability sampling of hospitals & patients. 2nd ed. Ann Arbor: Health Administration Press, 1975:15.

13. Holloway DC, Holton J, Goldberg GA, Restuccia JD. Development of hospital levels of care criteria. Health Care Management Review 1976;1:61.

14. Fuhrer R. A level of care index: development & evaluation of a computer aided utilization review selector method. Unpublished dissertation, Section on Medical Information Science, Division of Ambulatory & Community Medicine, University of California, San Francisco, June 1977.

15. Gertman PM, Bucher BM. Inappropriate hospital bed days and their relationship to length of stay parameters. Unpublished paper presented at the Medical Care Section, 99th Annual Meeting of the American Public Health Association, Minneapolis, October 11, 1969.

16. Neter J, Wasserman W. Applied linear statistical models. Homewood, IL: Irwin, 1974.

17. Commission on Professional and Hospital Activities. Length of stay in PAS hospitals, United States, Western Region, 1975. Ann Arbor, 1976.

18. Health Care Financing Administration. PSROs 1978 Program Evaluation. Office of Policy, Planning, and Research, Department of Health, Education and Welfare, Pub. No. HCFA-03000, January 1979.

19. Blau PM, Scott WR. Formal organizations: a comparative approach. San Francisco: Chandler, 1962.

20. Sanazaro PJ, Worth RM. Concurrent quality assurance in hospital care: report of a study by private initiative in PSRO. N Engl J Med 1978;298:1171.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Medical Information Systems: Assessing Impact in the Areas of Hypertension, Obesity and Renal Disease

JAMES L. ROGERS, PH.D.,* OLGA M. HARING, M.D.,† PAUL M. WORTMAN, PH.D.‡
ROYCE A. WATSON, M.B.A.,§ AND JOHN P. GOETZ, B.A.†

This study describes the influence of a computerized medical record summary system in three disease areas (hypertension, obesity and renal disease) observed in the course of a controlled, randomized and prospective study of 479 Northwestern University Clinic patients. Experimental patients, who had available automated record summaries, and control patients, who had available only the manual record, were compared on several medical tests and procedures whose yearly occurrence was considered good medical practice for this patient population, and were compared as well on several measures of outcome of medical care. Evidence suggesting better care and outcome of care among patients with computerized record summaries available is presented.

THERE IS CURRENTLY GREAT INTEREST in computer-automated medical records, and technical descriptions of the various medical information systems (MIS) and their capabilities are available.[1–16] It appears that computerized records are accepted by physicians, improve access to information, facilitate patient management and research and provide educational opportunities.[5,9,10,14,17–23]

Relatively few studies examine the impact of MIS on the actual practice and outcome of patient care. Bulpitt et al.[18] found no difference between hypertensive patients with and without automated record summaries on blood pressure measurements, dropout rates and the frequency of most laboratory investigations. McDonald[20–21] found that automation leads to fewer physician errors and that these errors occur in the first place due to the physicians' limited capacity to process information rather than to deficient medical reasoning. Finally, Rogers and Haring[24] found that although no relationship exists between the use of automated record summaries and the incidence of hospitalization, the length of hospitalization per visit declines by the second year after automated records are adopted.

Clearly, the available data are insufficient to allow the formulation of general conclusions, thus underscoring the need for objective assessments of the impact of MIS on patient care over a variety of disease areas and computer systems. The

---

Address for reprints: James L. Rogers, Ph.D., Department of Psychology, Wheaton College, Wheaton, IL 60187.

The major support for this project was provided by Grant Number HS 02649 from the National Center for Health Services Research, HRA. Initial data collection and analysis were made possible by DHEW Grant Number H500674-04 and USPAS Grant Number RR 05370 (NIH).

From the Department of Psychology, Wheaton College,* the Northwestern University Medical School,*† the School of Public Health, University of Michigan,‡ and the National Center for Health Services Research.§

0025-7079/82/0100/0063/$01.10 © J. B. Lippincott Co.

purpose of this article is to describe the influence of one MIS in three disease areas (hypertension, obesity and renal disease) observed in the course of a controlled, randomized and prospective study of 479 Northwestern University Clinic patients. Experimental and control patients are compared on several medical tests and procedures whose yearly occurrence was considered essential to good medical practice for this patient population and are compared as well on several measures of outcome of medical care.

To our knowledge, two of the disease areas under study, renal disease and obesity, have not yet been the subject of previous investigations. The third area, hypertension, has been investigated (see Bulpitt et al.[18]), but the present study covers a longer time frame, involves a different MIS and incorporates a more sensitive statistical analysis.

Methods

Computer System

We have developed at the Cardiac, Pulmonary and Renal Clinics of the Northwestern University a computerized medical record system (NUCRSS) to provide physicians with concise and current information on patients' problems, to identify omissions in recording of observations and treatment recommendations, to show ordered procedures that were not carried out, to record deficiencies in medical reasoning and, most importantly, to recommend corrective actions according to selected criteria. These criteria of "good care" were established by consensus of the clinicians providing care at our university.

Study Design

A prospective, randomized study was undertaken to evaluate the effect of the record summary on the practice and outcome of care. Patients were eligible for the study if they had attended the Cardiac, Pulmonary and Renal Clinics (CPRC) for more than six months and were not pregnant or under psychiatric care. From 1,200 eligible patients, 484 were randomly selected and assigned to either an experimental or control group. Five patients withdrew from the study before it began, leaving 241 experimental and 238 control patients. In the experimental group, patients had available a computer printout of a current NUCRSS summary in addition to the traditional medical record, while the control group had available only the handwritten, traditional medical record.

Physicians participating in the study were randomly divided into three groups: 1) those who were to see only patients with automated records available; 2) those who were to see patients without automated records; and 3) those whose patient load was approximately half with and half without automated records.

Data Base

One and two years after the entry of patients into the study, blind retrospective chart reviews were done for both experimental and control patients by trained personnel using a standardized evaluation form designed by us. Briefly, the resultant data base consisted of 1) items related to the utilization of services and to the overall quality of care (e.g., number of clinic visits, provider I.D., yearly routine physical examinations, recording of signs, occurrence and reason for hospitalization, etc.); 2) more detailed information, such as the presence or absence of recommended laboratory examinations for patients who had one or more of the following tracer diseases: hypertension, organic heart disease, renal disease, pulmonary disease, obesity and diabetes; and 3) answers to a questionnaire concerning patients' views on their own health and on the care received.

Standards of care were discussed and agreed upon by members of the Department of Medicine, relying upon various criteria. In the suggestions and comments of the record summary, physicians were reminded by the computer to adhere to these prescribed standards of care. For instance, if there was no evidence that serum potassium was checked periodically in patients on diuretic treatment, this was recommended; if potassium was low, appropriate medication was suggested; if patients with abnormal cultures were not treated, the suggestion was made to do so.

A description of the evaluation structure that was utilized to reflect the adequacy of care, along with a more comprehensive discussion of the design and data that were collected, has been published elsewhere.[25]

Data Presentation and Statistics

The data are presented in three parts in the results section that follows. First, dropout rates, baseline parameters and clinic attendance are compared across conditions. Second, selected medical tests and procedures whose yearly occurrence was considered by our physicians to be essential to the ongoing care of patients with chronic hypertension, obesity and renal disease were compared across conditions. These tests and procedures, which were monitored by the computer, are the examination of renal function (i.e., blood urea nitrogen, creatinine or creatinine clearance), the examination of potassium, the funduscopic examination and the intravenous pyelogram for hypertensive patients. For patients with renal disease we have selected the examination of renal function (as above), the urine analysis and the urine culture. Finally, whether or not a diet was given or reviewed was compared across conditions for obese patients. Chi-square statistics were used to determine whether differences across conditions in the incidence of items noted here were statistically reliable.

In each of the three disease areas, hypertension, obesity and renal disease, at least one indicator of outcome of care was collected and compared across conditions. Systolic and diastolic blood pressure measurements were recorded for hypertensive patients; the number of pounds overweight was periodically collected for obese patients; and the outcomes (normal or abnormal) during year 2 of the study for the renal function examination, the urine analysis and the urine culture were collected for patients with renal disease. The analysis of variance and the analysis of covariance were used to compare the experimental and control conditions on blood pressure and weight measurements, while chi-square statistics were used to compare patients with renal disease across conditions on the various laboratory examination results.

It is possible to compare all dependent variables across conditions by examining the various tables that follow. Values of statistics and the corresponding p levels are reported in table footnotes. Throughout the text, a p value of 0.05 or less is considered by the authors to be statistically reliable.

Results

Comparisons between the experimental and control groups are presented for 1) dropout rates, baseline parameters and clinic attendance; 2) selected medical care events that were considered a priori to reflect good medical practice for this patient population; and 3) several measurements of outcome of care.

Dropout Rates, Baseline Parameters and Clinic Attendance

Table 1 presents the number and per cent of patients in each of the three disease areas under investigation, along with the number and per cent of patients in each area who died by the end of the two-year study period or were transferred to another clinic, moved or left for unknown reasons. These figures indicate that differential dropout rates across conditions due to death or other reasons were not present.

Random assignment of subjects to conditions was conducted to equate by chance the conditions on factors other than the

TABLE 1. Patient Dropout Rates for Three Disease Areas

Disease Area/ Condition	No. of Patients with Disease (%)*	No. of Patients Dead by End of Study (%)†	Transferred to Another Clinic, Moved or Left for Unknown Reasons by End of Study (%)†
Hypertension			
Experimental	197 (81.7)	28 (14.2)†	28 (14.2)
Control	185 (77.7)	33 (17.8)	23 (12.4)
Obesity‡			
Experimental	68 (28.2)	1 (1.5)	5 (7.4)
Control	79 (33.2)	7 (8.6)	10 (12.7)
Renal Disease			
Experimental	64 (26.6)	10 (15.6)	14 (21.9)
Control	45 (18.9)	10 (22.2)	4 (8.9)

Notes: When the experimental and control conditions were compared on the above dropout rates, in no instance was the difference statistically reliable ($p > 0.05$).

* Percentages in parentheses based on total number of experimental (241) or control (238) patients in the study.

† Percentages in parentheses based on total number of experimental or control patients noted in column 1.

‡ Patients whose weight exceeded 20 per cent of their ideal wieght were classified as obese.

presence or absence of the MIS. To check for *random* differences at baseline, the following baseline parameters were compared: age; sex; race; fee schedule; prior clinic attendance in months; prior medical chart weight; the prior presence or absence of a) hypertension, b) renal disease, c) obesity, d) diabetes, e) organic heart disease and f) pulmonary disease; and the total number of diseases a through f present in each patient. As would be anticipated due to the use of random assignment, most differences between conditions on these variables were small. However, in a few instances, chance differences across conditions in baseline parameters were sufficiently large to warrant consideration as possible rival explanations for any differences on the dependent variables under study that otherwise would logically be attributed to the presence or absence of the MIS.

Baseline parameters judged by us to be potentially important are presented in Table 2, which indicates that 1) there were more hypertensive males in the control group than in the experimental group; 2) diabetes as a concomitant disease among obese patients was more prevalent in the experimental than in the control group; and 3) for all three disease areas, patients in the control conditions had been attending the Cardiac, Pulmonary and Renal Clinics prior to the beginning of the study for a longer average time than had the experimental patients. The implication of these baseline discrepancies for the interpretation of differences found on the dependent variables under study (see Tables 3 and 4) will be discussed throughout the following sections.

As a final comparison, the two conditions were contrasted on clinic attendance during year 1 and year 2 of the study. There was little difference across conditions, with both groups averaging between four and six visits per year per patient.

Medical Care Events

The variables selected to reflect important medical care events for this patient population are presented for each disease area in Table 3. Distributions indicating whether these events occurred during year 1 only, year 2 only, both years or not at all are provided for both the experimental and control conditions.

Unless otherwise noted, when differences between conditions on variables in Table 3 are cited in the following text as existing, it can be assumed that they are statistically reliable ($p < 0.05$). As noted earlier, the exact value of statistics and their corresponding p levels are presented in table footnotes, a procedure adopted throughout this and the following section.

Hypertension. Examination of the renal function and of potassium occurred more frequently during both years of the study in the experimental group. Failure to perform these examinations at all during the study period was more frequently found among the control patients. When these examinations occurred only once in the two-year study period, there was some tendency for the examinations to be done during year 1 for the experimental patients but during year 2 for the controls. There was little difference in distributions across conditions for either the funduscopic examination or the intravenous pyelogram.

Obesity. Regardless of whether the data are combined over sex or independently examined by sex, failure to give or review a diet at any time during the two-year study period clearly occurred less often among experimental patients (see Table 3).

Renal Disease. There were fewer experimental patients who had not had a urine analysis performed either year, while more experimental patients had tests performed both years (see Table 3). There were generally more failures to examine the renal function in the control condition, but the overall difference was not statistically significant. Likewise, the difference across conditions in the distribution of urine cultures was not statistically reliable.

TABLE 2. Baseline Parameters With Large Differences Across the Experimental and Control Groups for Patients With Hypertension, Obesity and Renal Disease*

Disease/Baseline Variable	Experimental	Control
Hypertension	N = 197	N = 185
Sex		
Male	80 (40.6%)	55 (29.7%)
Female	117 (59.9%)	130 (70.3%)
Prior clinic attendance (mean months)	80.9 ± 8.9†	104.1 ± 12.3
Obesity	N = 68	N = 79
Incidence of diabetes as a concomitant disease	30 (44.1%)	19 (24.1%)
Prior clinic attendance (mean months)	83.8 ± 15.6	107.2 ± 19.8
Renal Disease	N = 64	N = 45
Prior clinic attendance (mean months)	78.2 ± 15.6	96.9 ± 21.3

* Thirteen baseline parameters were compared across the experimental and control conditions: age, sex, race, fee schedule, prior clinic attendance in months, prior medical chart weight, the prior presence or absence of hypertension, renal disease, obesity, diabetes, organic heart disease and pulmonary disease, and the total number of these diseases present. The above differences were large enough to warrant further consideration.

† 95 per cent confidence intervals reported with means.

Baseline Differences. Within each disease area, various distributions for the dependent variables listed in Table 3 were cross classified along the corresponding discrepant baseline parameter(s) noted in Table 2. This was done to determine whether or not the differences cited above would remain unaltered after "controlling" for discrepant baseline parameters. In no instance was there a substantive change in the findings already reported for the dependent variables. Because no additional insight would be provided by displaying the multidimensional cross tabulations resulting from the incorporation of baseline parameters into the analysis, these figures have been omitted.

TABLE 3. Dependent Variables Reflecting Medical Care Events in Three Disease Areas for Experimental and Control Patients

Disease Area/ Procedure or Test	Done Year 1*		Done Year 2*		Done Both Years*		Not Done*	
	Experimental	Control	Experimental	Control	Experimental	Control	Experimental	Control
Hypertension								
Renal function examination†	44 (22.3)	38 (20.5)	18 (09.1)	26 (14.1)	120 (60.9)	93 (50.3)	15 (07.6)	28 (15.1)
Potassium	46 (23.3)	38 (20.5)	20 (10.2)	24 (13.0)	119 (60.4)	97 (52.5)	12 (06.1)	26 (14.1)
Funduscopic	9 (09.5)	6 (03.2)	119 (59.8)	100 (52.6)	14 (07.0)	9 (04.7)	55 (27.9)	70 (37.8)
Intravenous pyelogram	13 (06.5)	13 (06.8)	45 (22.6)	60 (31.6)	78 (39.2)	59 (31.1)	61 (31.0)	53 (28.6)
Obesity								
Number of diets given or reviewed combined over sex	11 (16.2)	9 (11.4)	20 (29.4)	16 (20.3)	23 (33.8)	16 (20.3)	14 (20.6)	38 (48.1)
Males	3 (15.0)	1 (06.7)	9 (45.0)	1 (06.7)	6 (30.0)	7 (46.7)	2 (10.0)	6 (40.0)
Females	8 (16.7)	8 (12.5)	11 (22.9)	15 (23.4)	17 (35.4)	9 (14.1)	12 (25.0)	32 (50.0)
Renal disease								
Renal function examination†	12 (18.8)	6 (13.3)	2 (03.1)	6 (13.3)	45 (70.3)	25 (55.6)	5 (07.8)	8 (17.8)
Urine Analysis	21 (32.8)	9 (20.0)	7 (10.9)	9 (20.0)	30 (46.9)	14 (31.1)	6 (09.4)	13 (28.9)
Urine Culture	31 (48.4)	27 (60.0)	6 (09.4)	5 (11.1)	16 (25.0)	9 (20.0)	11 (17.2)	4 (08.9)

Notes: Statistically significant differences across the experimental and control conditions were found for the examination of renal function (χ^2 [3] = 8.9, p = 0.03) and potassium (χ^2 [3] = 8.2, p = 0.042) among hypertensive patients; number of diets given or renewed among obese females (χ^2 [3] = 10.1, p = 0.018) and all obese patients combined over sex (χ^2 [3] = 12.1, p = 0.007); and for urine cultures (χ^2 [3] = 10.5, p = 0.015) among patients with renal disease. All other comparisons resulted in differences that were not statistically reliable ($p > 0.05$).

* Percentages in parentheses.

† Blood urea nitrogen, creatinine or creatinine clearance.

TABLE 4. Mean Blood Pressure for Hypertensive Experimental and Control Patients

	Unadjusted Means		Adjusted Means[*]	
	Experimental	Control	Experimental	Control
Systolic Blood Pressure				
At beginning of study				
Males	147.7 ± 4.9[†] (73)[‡]	149.1 ± 6.4 (53)	—	—
Females	149.9 ± 4.2 (110)	153.3 ± 4.8 (123)	—	—
After 10–15 months				
Males	145.5 ± 5.6 (65)	152.7 ± 6.7 (44)	147.7	151.8
Females	148.1 ± 4.8 (103)	147.9 ± 4.5 (117)	148.6	146.5
After 22–24 months				
Males	145.0 ± 8.4 (39)	148.2 ± 9.0 (25)	144.5	146.8
Females	146.2 ± 5.5 (80)	147.0 ± 5.4 (85)	146.9	147.0
Diastolic Blood Pressure				
At beginning of study				
Males	94.9 ± 3.2 (73)	93.4 ± 4.0 (53)	—	—
Females	93.0 ± 2.4 (109)[§]	91.9 ± 2.6 (122)[§]	—	—
After 10–15 months				
Males	92.4 ± 3.2 (65)	91.7 ± 4.4 (44)	91.5	91.4
Females	91.9 ± 2.8 (102)	90.4 ± 2.6 (117)	91.7	91.3
After 22–24 months				
Males	91.6 ± 3.9 (39)	94.4 ± 7.0 (25)	90.1	94.0
Females	91.2 ± 3.1 (80)	89.6 ± 2.8 (85)	91.3	90.1

Notes: For both systolic and diastolic blood pressure, the main effects of the treatment and of sex, as well as the treatment by sex interactions, were not statistically significant for either the unadjusted means using the analysis of variance or the adjusted means using the analysis of covariance.

[*] Covariance adjustment for blood pressure at the beginning of the study, prior time in cardiac-pulmonary-renal clinics and patient age.

[†] 95 per cent confidence intervals reported with unadjusted means.

[‡] Parentheses contain numbers of patients, which differ over time periods because not all patients visited the clinic during the 10–15 month and 22–24 month periods.

[§] In two instances diastolic blood pressure was not recorded, even though systolic blood pressure was available.

Patient Outcome

Hypertension. Table 4 contains the average systolic and diastolic blood pressures (measured at the beginning of the study) for patients with hypertension. Also included are averages that have been covariance adjusted for blood pressure value at the beginning of the study as well as for the two discrepant baseline parameters noted earlier, length of time spent in the Cardiac, Pulmonary and Renal clinics before the onset of the study and patient age. Both unadjusted and adjusted mean values have been computed for the time periods of 10–15 months and 22–24 months after the study had begun. Unadjusted mean values present at the beginning of the study are also given in Table 4. Use of the MIS, as evidenced by the lack of variability among means in Table 4, had no effect on systolic or diastolic blood pressure for either males or females at any point in time during the study. Details of these comparisons are found in Table 4.

Obesity. Each patient's ideal weight was subtracted from his or her weight at the beginning of the study, during the 10–15 month period and during the 22–24 month period. These differences in scores reflect "pounds overweight." In comparing the experimental and control groups, it was then necessary to adjust pounds overweight taken after the study began by the amount that the patient was overweight at the beginning of the study as well as by the patient's body size (i.e., ideal weight). The analysis of covariance was used to adjust for these two factors and for the discrepant baseline parameters noted earlier—prior attendance at the Cardiac, Pulmonary and Renal clinics and the incidence of diabetes as a concomitant disease. Additionally, the total number of other concomitant diseases each obese patient had was entered as a covariate (see footnote, Table 2). The mean unadjusted and the mean adjusted pounds overweight for males and females during the 10–15 month period and during the 22–24 month period after the study began are presented in Table 5. Also, the means for unadjusted pounds overweight at the beginning of the study are presented for the various groups.

Examination of Table 5 reveals that females are more overweight than males in both the experimental and control conditions. More importantly, by 22–24 months after the study began, females in the experimental group were less overweight than female controls and males in the experimental group were considerably less overweight than male controls.

Renal Disease. As measures of outcome of care for patients with renal disease, Table 6 contains the results of renal function examinations (i.e., blood urea nitrogen, creatinine or creatinine clearance), urine analyses and urine cultures that were given during year 2 of the study. More normal test outcomes were present among experimental patients for all three of these procedures, although the difference between conditions was statisically reliable only for the urine analyses and the urine cultures.

Newly Classified Patients. Over the course of the two-year study period, 14 (5.8 per cent) of the 241 experimental patients and 10 (4.2 per cent) of the 238 control patients were newly classified as hypertensive, a difference not statistically significant ($Z = 0.6$, $p > 0.05$). Corresponding figures for newly classified obese patients are 13 (5.4 per cent) experimental and 20 (8.4 per cent) control patients ($Z = 1.1$, $p > 0.05$). The findings previously reported for hypertensive and obese patients held for newly classified patients as well. That is, there was little difference between experimental and control patients on blood pressure measurements, while the 10–15 month period experimental patients tended to be less overweight than controls (28.2 versus 36.7 pounds for males, and 22.1 versus 42.6 pounds for females, $F[1, 29] = 4.8$, $p < 0.04$). There were no experimental or control patients that were newly classified as having renal disease during the study period.

Discussion

Generally, the experimental patients underwent more laboratory procedures and/or were given (or had reviewed) more diets than the control group. These events were considered a priori as essential to good, routine care for this population of patients. Of eight procedures examined among patients in three disease areas, six were performed more frequently among experimental than among control patients (see Table 3). Of these six differences, four were statistically significant. Further, obese patients in the experimental groups

Table 5. Mean Pounds Overweight for Obese Experimental and Control Patients

	Unadjusted Means		Adjusted Means*	
	Experimental	Control	Experimental	Control
At beginning of study				
Males	50.3 ± 6.1† (15)‡	58.2 ± 13.4 (11)	—	—
Females	54.5 ± 8.3 (40)	55.0 ± 6.2 (48)	—	—
After 10–15 months				
Males	44.8 ± 8.4 (15)	56.5 ± 16.4 (11)	45.6	48.6
Females	51.5 ± 9.1 (40)	54.5 ± 7.3 (46)	52.3	55.3
After 22–24 months				
Males	34.8 ± 8.6 (11)	59.5 ± 19.0 (9)	39.3	52.2
Females	52.0 ± 10.7 (35)	54.8 ± 7.9 (35)	51.5	55.8

Notes: Analysis of covariance shows that during the 10–15 month period, neither the treatment by sex interaction nor the main effect of the treatment was statistically significant (F [1,100] = 0.55, p = 0.45, and F [1,103] = 2.4, p = 0.12, respectively) but that the main effect for sex was significant (F [1,103] = 4.1, p = 0.045). For the 22–24 month period, the treatment by sex interaction and the main effect of sex were not statistically significant (F [1,81] = 1.8, p = 0.18, and F [1,81] = 3.2, p = 0.08, respectively), while the main effect of the treatment was significant (F [1,81] = 5.4, p = 0.023).

* Covariance adjustment for pounds overweight at the beginning of the study, ideal weight, prior time in cardiac-pulmonary-renal clinics, presence or absence of diabetes as a concomitant disease and the total number of other concomitant diseases.

† 95 per cent confidence intervals reported with unadjusted means.

‡ Parentheses contain numbers of patients, which differ over time periods because not all patients visited the clinic during both the 10–15 month and 22–24 month periods.

were less overweight, while experimental patients with renal disease had better results on year 2 urine analyses and urine cultures.

The traditional record used in the clinics contained on the average 1½ pounds of notes, laboratory test outcomes, diagnoses and other information entered over long periods of time by different physicians. This state of affairs no doubt jeopardized continuity of care, especially since patients were not necessarily seen by the same physician from visit to visit. The present MIS, in addition to condensing the medical information to items that were current, legible and relevant, provided physicians with warnings and reminders concerning good medical practices. For this reason it was expected that the MIS would have a positive effect on medical care, and it was the sole purpose of the present experiment to determine whether this occurred by examining differences between conditions on selected dependent variables. Not knowing whether any differences would be found (or what the nature of the differences would be), this study was not designed to examine the specific mechanisms underlying any effects that the MIS might have. However, in order to attribute any effects that might be found to the presence or absence of the MIS rather than to extraneous differences between conditions on uncontrolled variables, we randomly assigned patients to conditions, randomly assigned physicians to conditions and analyzed the dependent variables controlling for the few chance baseline differences that appeared large enough to present a potential threat to the internal validity of this experiment. Further, the presence of differential dropout rates and differential clinic attendance during the study was ruled out by comparing these variables across conditions. In light of these procedures and facts, the differences between the experimental and control groups on the dependent variables cited here seem reliable, and these differences suggest that the MIS has had a positive effect on patient care.

As noted, questions remain concerning the mechanisms by which changes occur, and clarifying studies in this area will be needed. To this end, our impressions of the underlying causes of some of the differences reported here may be helpful to other investigators or to those considering the adoption of an MIS. The vast differences that exist in types of systems, clinical environments and needs, patient populations and even criteria of good care must of course be kept in mind.

Results in the area of obesity management seem the clearest. Experimental patients either had new diets assigned or old diets reviewed more frequently than control patients. Correspondingly, experimental patients evidenced greater weight loss on the average than did control patients. The NUCRSS keeps track of weight loss progress and reminds physicians to review or change diets. We hypothesize that this resulted in greater awareness on the part of physicians and patients, which in turn promoted weight loss.

A secondary finding of some interest is that more progress in weight loss was made by males than females, and that even though experimental patients were less obese than control patients, after 22–24 months they remain considerably overweight (39.3 average adjusted pounds for males and 51.3 average adjusted pounds for females). This is not surprising in that most weight loss programs reported in the literature, regardless of orientation, have been only marginally successful.[26]

For experimental patients with renal disease, it is possible that because more urine analyses were ordered during the first year of the study for them than for the control group, problems indicated by these tests could have been corrected by the second year. Additionally, the increased awareness of test results that would be expected when the MIS was used should have the overall effect of inducing appro-

TABLE 6. Results of Three Laboratory Examinations for Patients with Renal Disease

Test/Result of Test	Experimental (N = 54)		Control (N = 33)	
	Done	Not Done	Done	Not Done
Renal function examination[*]	47 (87.0)[†]	7 (13.0)	31 (93.9)	2 (06.1)
Normal result	17 (36.2)[‡]		7 (22.6)	
Abnormal result	30 (63.8)		24 (77.4)	
Urine analysis	37 (68.5)	17 (31.5)	23 (69.7)	10 (30.3)
Normal result	17 (45.9)		1 (04.3)	
Abnormal result	20 (54.1)		22 (95.7)	
Urine culture	22 (40.7)	32 (59.3)	14 (42.4)	19 (57.6)
Normal result	15 (68.2)		5 (35.7)	
Abnormal result	7 (31.8)		9 (64.3)	

Notes: When the percentages of normal results were compared across the experimental and control conditions, the following one-tailed Z statistics emerge: for the Urine Analysis, $Z = 3.40$, $p = 0.0003$; for the Urine Culture, $Z = 1.91$, $p = 0.028$; and for the Renal Function Examination, $Z = 1.27$, $p = 0.10$.

[*] Blood urea nitrogen, creatinine or creatinine clearance.

[†] Percentages in parentheses based on number of patients in the study at the end of year 1 (i.e., N = 54 or N = 33).

[‡] Percentages in parentheses for outcomes based on total number of tests done during year 2.

priate medical care and thus improving results on the urine analyses and urine cultures the following year. Unfortunately, first-year results of laboratory investigations were not available to us, and this hypothesis cannot therefore be fully tested. It is not unreasonable from a medical point of view, however, to believe that abnormalities detected in the urine analysis or urine culture could often be cleared up by the second year once they were detected. On the other hand, abnormalities found in the renal examination usually would be much more difficult to eliminate. Correspondingly, a large difference between conditions in results would not be expected and, in fact, was not found for the outcomes of renal function examinations.

The data do not suggest a simple causal interpretation in the area of hypertension. In the present study, the evaluation of renal function and the examination of potassium occurred more frequently in the experimental than in the control groups. In the Bulpitt et al. study cited earlier, the incidence of the examination of potassium was reviewed, but, unlike the present study, no statistically reliable difference was found. These investigators also examined the incidence of blood glucose examinations and found a higher incidence of this procedure for experimental patients. Although the above differences suggest that important medical events for the ongoing care of hypertensive patients may occur more routinely when an MIS is available, apparently such changes have not resulted in improved outcome of care as indexed by blood pressure values. No difference on blood pressure was found by Bulpitt et al. or by us. We had expected that better information, through such means as heightening physician awareness of medication regimens or facilitating patient compliance due to better physician–patient communication, would result in reduced blood pressure. Whatever the reasons for continued high blood pressure in both groups (e.g., noncompliance of patients with medication regimens, only partially effective medications, etc.), the computer summary has not improved (or worsened) the situation.

In this study we used simple randomization (rather than stratifying the sample before randomization), followed by the statistical control of the occasional baseline differences that emerged by chance. If the experiment could be redone, it would be best to stratify the sample along the three baseline dimensions found to differ (sex of diabetic patients, presence of diabetes for obese patients and prior clinic attendance for all patients), but there was, of course, no way of knowing before the random assignment of patients that these three variables, rather than some of the other 10 baseline variables that were also retrospectively checked, would be different across conditions. Probably the best procedure for future MIS evaluations would be to select a few important baseline variables for stratification purposes to assure equality across conditions on at least these variables, and then to adjust statistically for other baseline variables as necessary, after retrospectively examining them for differences between conditions.

As a concluding comment, we note that the simmering debate presently underway over process versus outcome measures of health care quality has some bearing on the present report.[8] For example, what is often considered "process" of care (the occurrence of important medical tests or diets) was improved among hypertensive and obese patients who had the MIS available. Yet, outcome of care was not improved among hypertensive patients (i.e., blood pressure remained high) and only relatively improved among obese patients (i.e., relative to the control groups, experimental patients were statistically less overweight, but they were still considerably overweight). Our position is that as medical knowledge advances, the relationship between process and outcome of care will improve. Therefore, the worth of broadly based technologies like MIS should be judged on the basis of how well they aid the occurrence of what is currently and locally considered good medical practice (i.e., good process of care), even if current practice is less effective in terms of outcome than would be hoped. Accordingly, the present system has made a positive impact on the occurrence of medical events considered important for the present patient population and should help to insure the occurrence of any new, more efffective procedures that might develop in the future.

References

1. Barnett GO. Computer-stored ambulatory record (COSTAR). Washington D.C., DHEW Pub. No. (HRA) 76-3145, 1976.

2. Beilin LJ, Bulpitt CJ, et al. Computer based hypertension clinical records: a co-operative study. Br Med J 1974;2:212.

3. Bradshaw-Smith JH. A computer record-keeping system for general practice. Br Med J 1976;1:1395.

4. Cobelli C, Salvan A. A medical record and a computer program for diagnosis of thyroid disease. Methods Inf Med 1975;14:126.

5. Grossman JH, Barnett GO, et al. An automated medical record system. JAMA 1973;124:1616.

6. Jerome HH, Barnett GO, et al. An automated medical record system. JAMA 1973;224:1616.

7. Kennedy F, Cleary JJ, et al. "Switch": a system producing a full hospital case history on computer. Lancet 1968;2:1230.

8. Levy RP, Cammarn MR, Smith MJ. Computer handling of Ambulatory clinical records. JAMA 1964;190:1033.

9. McDonald CJ, Murray R, et al. A computer-based record and clinical monitoring system for ambulatory care. Am J Public Health 1977;67:240.

10. Mellner C, Selander H, Wolodarski J. The computerized problem-oriented medical record at Karolinska Hospital—format and function, user's acceptance and patient attitude to questionnaire. Methods Inf Med 1976;15:11.

11. NCHSR Research Digest Series: Automation of the problem-oriented medical record. Washington, D.C., DHEW Pub. No. (HRA) 77-3177, 1977.

12. Schenthal JE, Sweeney JW, Nettleton WJ. Clinical applications of large scale electronic data processing apparatus I. JAMA 1960;173:6.

13. Schenthal JE, Sweeney JW, Nettleton WJ, Yoder RD. Clinical applications of large scale electronic data processing apparatus III. JAMA 1963;186:101.

14. Schmidt EC, Schall DW, Morrison CC. Computerized problem-oriented medical record for ambulatory practice. Med Care 1974;12:316.

15. Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer based medical history system. N Engl J Med 1966;274:194.

16. Wickham JEA, Charlton CAC, et al. A computer-based record and organization system for a department of urology. Br J Urol 1975;47:345.

17. Battele Columbus Laboratories. Evaluation of medical information system in a community hospital. Washington, D.C., DHEW Pub. No (HRA) 76-3144, 1976.

18. Bulpitt CJ, Beilin LJ, et al. Randomised controlled trial of computer-held medical records in hypertensive patients. Br Med J 1976;1:677.

19. Ehrenhaft PM, Corcoran S, Harwood EA, West H. Policy implications of medical information systems. Washington, D.C.: Office of Technology Assessment, 1977.

20. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. N Engl Med 1976;295:1351.

21. McDonald CJ. Use of a computer to detect and respond to clinical events: its effect on clinical behavior. Ann Intern Med 1976;84:162.

22. Starfield B, Simborg D, Johns C, Horn S. Coordination of care and its relationship to continuity and medical records. Med Care 1977;15:929.

23. Zielstorff RD, Roglieri JL, et al. Experience with a computer-based medical record for nurse practitioners in ambulatory care. Comput Biomed Res 1977;10:61.

24. Rogers JL, Haring OM. The impact of a computerized medical record summary system on incidence and length of hospitalization. Med Care 1979;17:618.

25. Lennie FS, Ametrano DA, Haring OM. Evaluation of an automated record summary for ambulatory care. Medinfo 1974;1:115.

26. Janda LH, Rimm DC. Covert sensitization in the treatment of obesity. J. Abnorm Psychol 1972;-80:37.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Assessing Medical Practices Through PSRO Cooperative Studies

An Evaluation of Cesarean Births in Nine PSRO Areas

WILLIAM F. JESSEE, M.D., C. W. NICKERSON, M.D., AND
WILLIAM S. GRANT, for the PSRO Coordinating Group*

Nine Professional Standards Review Organizations jointly conducted an areawide medical care evaluation study of Cesarean births. Of 5,002 Cesarean births studied in 158 hospitals, only 1.1 per cent were judged to be unjustified procedures by local physician peer review. Although there was substantial geographic variation in rates of Cesarean birth (11.0%–18.3 per cent), no statistically significant pattern of variation was identified. Maternal outcomes and infant outcomes were generally good. This study demonstrates the potential for use of PSRO data in studying major issues in health care, as well as the practical constraints of such studies.

THE PROFESSIONAL STANDARDS REVIEW Organization (PSRO) program was originally developed by the federal government to assure that health care services funded under the Social Security Act are medically necessary, appropriate and of acceptable quality.[1] In the course of carrying out these statutory duties, PSROs collect a large variety of health care data that are of potential value for increasing understanding of the forces that affect health services utilization and health care quality. Despite the attention that has been devoted to the educational role of PSROs in improving health care practice,[2] their potential as resources for evaluating contemporary community medical practices has not been adequately explored. This study represents a first attempt by a group of PSROs to pool data on an issue of substantial controversy in medical practice in an effort to define further its proper role.

Considerable attention has been focused during recent years on the rising rate of Cesarean birth in the United States.[3–6] Critics of the rising rates have pointed to the higher costs, associated morbidity and potential for late adverse effects for mother

---

From the Department of Health Administration, University of North Carolina School of Public Health, and the Colonial Virginia Foundation for Medical Care.

* The participating PSROs and their Coordinating Group representatives were: Bronx Medical Services (NY) (Seymour Cohen, M.D., Veronica M. Whalen); Colonial Virginia Foundation for Medical Care (C. W. Nickerson, M.D., William S. Grant, Sandra Latham); Essex Physicians Review Organization (NJ) (James T. Carroll, M.D., Jane Majcher); Iowa Foundation for Medical Care (Charles Driscoll, M.D., Jennifer Cofer); New Mexico PSRO (Phillip Kamps, M.D., John B. Cappes, M.D., Jim Reese); Northern Illinois PSRO (Bruce E. Greenfield, M.D., John Leonard, M.D., Donna Furman); Physicians Peer Review Organization (OH) (Lawrence H. Malm, M.D., Maria Hewett); Region Six Peer Review Organization (OH) (Robert H. Baxter, M.D., Faith A. Lower); Santa Clara Valley PSRO (CA) (William F. Jessee, M.D., Emily Wieczorek).

Address reprint requests to: William F. Jessee, M.D., University of North Carolina, Department of Health Administration, 266 Rosenau Building 201/H, Chapel Hill, NC 27514.

0025-7079/82/0100/0075/$01.00 © J. B. Lippincott Co.

TABLE 1. Screening Criteria for Evaluation of Primary Caesarean Sections

Criterion Number	Elements	Std.	Exceptions	Definitions and Retrieval Sources
1	*Justification for Surgery* One or more of the following:			
	A. Previous uterine or vaginal surgery	100%	None	*Uterine surgery*—myomectomy, metroplasty, hysterotomy, Strassman reunification procedure, repair of uterine perforation due to trauma or repair of incompetent cervix by Shirodkar or Lash procedure. *Vaginal Surgery*—anterior vaginal repair, repair of cystocele, posterior vaginal repair, repair of rectocele, vaginal reconstruction, or plastic surgery or repair of fistula (i.e., vesico-vaginal fistula repair, rectovaginal fistula repair, cervicouterine fistula repair, or uterine-vaginal or other fistula) *Do Not Accept* dilation and curettage (D&C), minor vaginal procedures, uterine suspension or other gynecological surgical procedures *Source:* History and physical, physician progress notes, operative notes, previous medical record if unit record
	OR			
	B. Failure to progress in labor & C-Section performed minimum 6 hours and maximum 18 hours after active phase of labor has begun	100%	None	*Failure to Progress*—in labor documented; may also be described as "failure to descend," "cervical dystocia" or "failure to dilate" *Active phase of labor*—dilation of 4 cm or greater *Source:* Labor record, operative notes, or physician progress notes
	OR			
	C. Fetal distress	100%	None	*Fetal distress*—confirmed by electronic fetal monitoring or scalp blood pH of less than 7.2 or by clinical assessment (auscultation, etc.) *Source:* Same as 1B.
	OR			
	D. Placental insufficiency	100%	None	*Placental insufficiency*—same as 1C and positive antepartum stress test (positive OCT—oxytocin challenge test) *Source:* Same as 1B and lab report
	OR			
	E. Maternal hemorrhage	100%	None	*Maternal hemorrhage*—abruptio placenta, placental abruption or placenta previa, vasco previa *Source:* Labor record, operative notes, physician progress notes or history
	OR			

	F. **Malpresentations in labor or with ruptured membranes**	**100%**	**None**	***Malpresentations*—breech, transverse lie, shoulder, anterior, face, brow** *Source:* Labor and delivery record, history, physician progress notes, operative notes
	OR			
	G. Herpes genitalis if membranes intact or ruptured less than 4 hours	100%	None	*Herpes genitalis*—documented by culture or cytology *Source:* History, physician progress notes, laboratory and pathology reports
	OR			
	H. Feto-pelvic disproportion in labor	100%	None	*Feto-pelvic Disproportion* (FPD)—cephalopelvic disproportion, hydrocephaly as evidenced in X-ray, ultrasound scan or physical examination *Source:* X-ray or ultrasound reports, physician progress notes, labor and delivery record, or operative report
	OR			
	I. Maternal insulin-dependent diabetes mellitus *with* fetal maturity studies	100%	None	*Insulin Dependent*—must be taking insulin by injection or syringe *Fetal Maturity Studies*— a. Amniocentesis AND b. L/S (lecithin–sphingomyelin) ratio within 2 weeks of procedure OR c. Ultrasound anytime during pregnancy *Source:* History, physician progress notes, laboratory, X-ray, or ultrasound reports
	OR			
	J. Prolapsed cord	100%	None	*Prolapsed cord*—premature expulsion of the umbilical cord in labor before the fetus is delivered *Source:* Physician progress notes, history, Operative notes or labor record
	Patient Outcome			
2	Maternal Mortality	0%	None	
3	Mortality of infant during hospitalization	0%	None	
	Length of Stay			
4	Minimum 3 days postoperative	100%	None	
5	Maximum 7 days postoperative	100%	None	
	Other			
6	Infant discharged with mother	100%	a. Transfer to another acute care facility for neonatal intensive care or transfer to in-house intensive care unit b. **Infant died** c. **Maternal complications**	

and child that are associated with Cesarean birth. Apologists have pointed out that the rising rates have not been accompanied by an increase in overall maternal mortality and that selective use of Cesarean birth in high-risk deliveries may improve the prognosis for the child. However, consensus on indications for Cesarean birth and approaches to its optimal utilization have not been forthcoming. We report here the results of a cooperative study of 5,002 Cesarean births performed in 158 hospitals during the period from July 1, 1977 through June 30, 1978. This study was performed by physicians and staff of nine PSROs as a demonstration of the role these organizations can play in an evaluation of important issues in medical practice.

Methods

This study was conducted as a joint medical care evaluation (MCE) study by nine participating PSROs. The participating PSROs were self-selected from a group of 30 who had initially been approached by one of the authors to solicit their voluntary participation in the study as a demonstration of the potential of such cooperative studies. The criteria for the study (Table 1) were set by consensus of the physician members of the coordinating group. Draft criteria based on a brief review of the obstetric literature were initially developed by one of the authors and circulated to each of the nine PSROs.

In each area, the input of local physicians performing Cesarean deliveries (both obstetricians and family physicians) was solicited concerning the study design and criteria. A variety of mechanisms were employed by individual PSROs in obtaining such input, ranging from referral to existing PSRO and hospital medical care evaluation (audit) committees, to requesting advice from local obstetric societies, to direct mailings to hospitals and physicians. Approximately two months after the draft criteria were distributed, a physician and a staff representative from each of the nine participating PSROs met in Chicago to finalize the criteria and study design. The criteria finally agreed upon represent the consensus of physicians engaged in the practice of obstetrics in the participating PSRO areas, as reflected by their representatives on the coordinating group. Criteria were set in four categories: justification for surgery (indications), maternal outcome, length of stay and infant outcome.

The objectives of the study were to evaluate regional variations in indications, length of stay and other parameters associated with the variable incidence of Cesarean birth; to evaluate regional differences in maternal complications; and to evaluate immediate neonatal outcomes.

In each of the participating hospitals, medical records of patients undergoing primary Cesarean birth were selected chronologically, starting with patients discharged on or before June 30, 1978. All patients undergoing the procedure during the one-year period preceding that date, to a maximum of 50, were included in the study. The limitation of the sample size to a maximum of 50 records is consistent with generally accepted procedures for medical audit and was imposed to minimize the burden of data collection in any individual hospital. If more than 50 primary Cesarean births occurred, the first 50 were chosen in reverse chronologic order of discharge. More than 50 Cesarean deliveries occurred in 77 (49 per cent) of the 158 participating hospitals. Repeat Cesarean births were excluded. Additional information items were collected in each of the 158 hospitals, including the total number of deliveries, primary and repeat Cesarean births during the year; hospital size, teaching status, urban/rural location and type of ownership; use of fetal monitoring in Cesarean births; frequency of concomitant tubal ligation; and frequency of absence of prenatal care in mothers delivered by Cesarean.

Data were submitted by the hospitals to each local PSRO and were aggregated by the coordinating group. Variations from the criteria were analyzed at the hospital level, and decisions were made by peer physicians in each institution concerning the justification of those variations. Each hospital, as is normal procedure in areawide MCE studies, used either members of its own audit committee or peer physician consultants from their staff to make judgments regarding justified variations. Plans for correction of deficiencies were also developed at the hospital or PSRO level. The aggregated national data were analyzed by PSRO area.

Results

In the nine areas under study, the incidence of primary Cesarean birth during the year studied was 9.9 per cent (range 7.2 to 12.8 per cent). When repeat Cesarean births are added, the overall rate rises to 14.7 per cent (range 11.0 to 18.3 per cent). Higher total Cesarean section rates were observed in the six PSRO areas that are predominantly urban (mean 16.1 per cent, range 14.1 to 18.3 per cent) than in the three predominantly rural areas (mean 12.1 per cent, range 11.0 to 14.8 per cent). However, these differences were not statistically significant ($p > 0.05$, df = 7). Similarly, the rates for primary Cesarean section were also higher in urban than in rural areas, but not at a statistically significant level (urban mean 11.0 per cent, range 9.7 to 12.8 per cent, rural mean 8.1 per cent, range 7.2 to 9.9 per cent; $p > 0.10$, df = 7). No other patterns of geographic difference in rates were discernible. The average length of hospital stay (ALOS) ranged from 5.1 days in PSRO area F to 8.6 days in area H. The two western areas exhibited the

TABLE 2. Indications for Cesarean Section Reported in 5,002 Births in Nine PSRO Areas (% of Total Cases Meeting Indication—Totals Exceed 100% Due to Presence of Multiple Indications in Some Patients)

Indication*	PSRO Area									All Areas
	A	B	C	D	E	F	G	H	I	
1. Feto-pelvic disproportion	52.8	41.6	39.8	41.4	47.7	33.0	48.6	28.2	29.0	40.4
2. Failure to progress in labor	25.7	25.2	26.5	38.8	38.1	22.0	28.3	30.9	27.0	30.6
3. Malpresentation	22.1	28.8	21.0	30.6	32.3	19.0	27.8	21.5	30.2	26.1
4. Fetal distress	17.2	16.4	18.8	14.6	14.2	15.3	13.4	34.1	16.5	17.0
5. Maternal hemorrhage	5.4	5.6	5.7	7.4	10.2	8.0	6.7	5.6	10.2	7.2
6. Previous uterine or vaginal surgery	2.5	2.0	1.5	2.4	1.5	1.8	0.2	2.4	3.4	2.0
7. Prolapsed cord	1.3	0	1.6	2.3	1.5	2.0	1.2	2.9	2.0	1.8
8. Placental insufficiency	2.0	0.8	0.4	0.9	0.5	2.2	1.8	1.2	0.5	1.2
9. Insulin-dependent diabetes mellitus	1.1	0.4	1.0	0.7	2.0	2.0	0.8	0.9	0.5	0.9
10. Herpes genitalis	0	0	0.8	0.1	0	1.5	1.4	0	0.7	0.5
11. Other justified	4.3	6.0	7.8	4.7	3.0	18.2	6.1	0.6	NA	6.1†
12. Unjustified	0.9	0.4	0.8	1.0	3.6	1.3	1.4	0.3	NA	1.1†
Total Number of Cases	557	250	618	1537	197	600	492	340	411	5002

* Definitions for each indication are shown in Table 1.

† Does not include Area I since peer analysis of cases not meeting screening criteria was not reported in this area.

TABLE 3. Maternal and Infant Outcomes and Complications Reported in 5,002 Cesarean Births in Nine PSRO Areas

PSRO Area	Cases Studied	Maternal Mortality (per 1,000)	Infant Mortality (per 1,000)	Maternal Complications(%)*	Infant Complications(%)*
A	557	0	10.8	12.7	9.0
B	250	0	4.0	9.6	3.6
C	618	0	21.0	10.0	4.2
D	1,537	0.65†	12.4	9.3	4.1
E	197	0	20.3	1.5	3.6
F	600	1.7†	20.0	4.5	9.8
G	492	0	14.2	7.9	2.8
H	340	0	13.3	13.2	15.0
I	411	0	17.2	22.9	2.2
All areas	5,002	0.41	15.3	10.2	5.8

* Maternal length of stay greater than seven days was used as a proxy measure of maternal complications. Infant discharge later than maternal discharge was used as a proxy measure of infant complications.

† One maternal death occurred in each area.

shortest ALOS (5.1 and 5.6 days), while the longest stays were in the northeastern areas (8.6 and 7.5 days) and in one of the north central areas (8.1 days).

The indications for Cesarean birth and the frequency of each indication are shown in Table 2. Totals exceed 100 per cent because of the presence of multiple indications in some patients. Maternal and infant outcomes in each area are presented in Table 3. Postoperative length of stay of greater than seven days was used as a proxy measure for maternal complications, and infant discharge after maternal discharge served as a proxy measure of infant complications. Table 4 displays data on prenatal care, fetal monitoring and concomitant tubal ligation among the 5,002 Cesarean births studied. Area H is notable in that it exhibited an unusually low incidence of prenatal care (78.5 per cent) and a very high rate of infant complications (15.0 per cent).

Discussion

This cooperative medical care evaluation study, or medical audit, demonstrates the potential of the PSRO program for

TABLE 4. Prenatal Care, Fetal Monitoring and Concomitant Tubal Ligation Reported in 5,002 Cesarean Births in Nine PSRO Areas

PSRO Area	Cases Studied	Prenatal Care Documented(%)	Fetal Monitoring Performed(%)	Concomitant Tubal Ligation Performed(%)
A	557	97.7	56.9	7.4
B	250	100.0	59.6	10.4
C	618	91.9	62.6	8.1
D	1,537	99.7	43.5	9.8
E	197	97.5	43.7	7.1
F	600	93.5	39.7	14.2
G	492	99.6	63.8	6.1
H	340	78.5	61.8	5.9
I	411	98.3	37.0	7.8
All areas	5,002	96.1	50.4	9.0

evaluating regional differences in health care delivery. The ability to collect data from hospitals that vary in size, location, ownership and teaching status allows analysis of community health practices in a way quite different from the usual university-based health care research program. Although the cooperative study reported here is clearly quite preliminary, its success in stimulating comparison of regional practice variations has led to plans for a more rigorously designed study of technological factors affecting Cesarean birth rates, using PSROs as the data collection intermediaries.

The design of this study followed that of conventional medical audits and therefore contains several biases that necessitate cautious interpretation of the data. For instance, the self-selection of the nine participating PSROs from a larger group of PSROs, all of which we felt might be interested, raises sampling questions. Further, as is the usual practice with PSRO areawide MCE studies, hospital participation was voluntary, introducing the possibility of additional sampling bias. Nonetheless, the geographic diversity of the PSROs (Table 5) was such that we consider the data at least representative of the range of regional variation, if not representative of the total U.S. experience. Further, more than 85 per cent of the hospitals performing Cesarean deliveries in the nine PSRO areas participated in the study.

Sampling in medical audits traditionally draws consecutive records, up to a preset maximum (in our study, 50 records was the maximum in any one hospital). Of the 158 hospitals, the "sample" studied represented the entire population of Cesarean births during the one-year period in 81 hospitals. In the remaining 77, the 50 consecutive records sampled represent a potential source of nonsystematic bias, which must be considered in interpreting the results.

Despite these caveats, our data indicate that the rate of Cesarean birth continues to rise in the United States. The overall rate reported by the 158 hospitals participating in this study (14.7 per cent of all deliveries) during the period from July 1, 1977 to June 30, 1978 exceeds both the 12.8 per cent figure reported for 1977 and the 13.9 per cent projection for 1978 cited by Marieskind.[6] Approximately one third of these procedures were repeat operations, a proportion comparable to that noted in the 1974 study of PAS (Professional Activities Study) hospitals reported by Lowe.[7] So long as the incidence of primary Cesarean birth continues to increase, the pool of women who will undergo repeat operation

TABLE 5. Characteristics of the Participating PSROs

PSRO	Primarily Urban	Primarily Rural	Statewide	Geographic Area*	No. Hospitals In Study	No. Patients In Study
A	X			NE	12	557
B	X			C	5	250
C	X			SE	15	618
D		X	X	C	70	1,537
E		X		C	9	197
F		X	X	W	15	600
G	X			W	10	492
H	X			NE	7	340
I	X			C	15	411
Totals	6	3	2	-	158	5,002

* NE = Northeastern U.S.; C = Central U.S.; SE = Southeastern U.S.; W = Western U.S.

for subsequent preganancies will also continue to increase. In the 1974 PAS study,[7] only 0.9 per cent of deliveries following previous Cesarean birth were vaginal. This policy of routine repeat procedures is in contrast with the estimated 50 per cent of deliveries following previous low-segment Cesarean birth that some authors feel may be safely performed vaginally.[6]

Among the 4,591 primary Cesarean sections in our study for which peer review of cases not meeting the screening criteria was performed, only 49 (1.1 per cent) were determined by local physician peers not to be justified. Elimination of this small proportion of cases would only lower the primary Cesarean birth rate from 9.9 to 9.8 per cent, an inconsequential change. Thus, it appears that the relatively high rates of Cesarean birth being reported cannot be described as "unnecessary" surgery by contemporary local or even national standards. Rather, the rise must, in large part, be attributed to changing indications for the procedure and to changing obstetric custom.

However, the extent to which some of the criteria used in our study can be validated by scientifically sound clinical data is of concern. For instance, several of the indications included in our screening criteria are quite subjective. "Failure to progress in labor" and "fetal distress," despite our attempts to define them, remain largely matters of the judgment of the attending obstetrician. There is little question that the criteria used represent consensus of the obstetric community and are thus empirically sound descriptions of current community standards. However, one major study that attempted to determine the scientific validity of criteria developed in this manner found disappointing results.[8] This requires that our finding of only 1.1 per cent of Cesarean births not justified by peer judgment be interpreted with caution. Until such time as more scientifically valid data concerning indications for Cesarean birth are developed, consensus criteria, despite their acknowledged deficiencies, will probably continue to be used in PSRO quality assessment activities. This is a serious weakness, and careful analysis of its policy implications is required, since such criteria are used in virtually all hospital quality assurance activities.

The issues of the influence of convenience, economics and the impact of obstetric intervention itself on the rate of Cesarean birth are speculative and are likely to remain so until they are adequately investigated. In our series, "failure to progress in labor" and "feto-pelvic disproportion" were the most frequently observed indications. Some would contend, however, that many patients who undergo operations for failure to progress are in fact the products of failed induction or augmentation of labor. Whether or not the risk of delay in delivery to mother or infant in such circumstances outweighs the risk of Cesarean birth is largely an unexamined issue. Is it necessary to induce labor in all cases of prematurely ruptured membranes (PROM), or is more conservative management indicated? Kappy et al.[9] report excellent results from nonintervention in 188 women with PROM under 37 weeks of gestation. Unfortunately, our study did not examine the relationship between PROM and Cesarean delivery. This issue, however, will be the subject of further investigation, as will the timing of the choice between Cesarean and vaginal delivery in mothers with PROM.

Oxytocin administration for induction or augmentation of labor may be a major factor in the increased rate of Cesarean births.[10] Difficulty in maintaining adequate control of contractions can produce uterine tetany, induce fetal distress, or lead to early fatigue and arrested labor; conversely, inadequate trials of oxytocin may lead to a premature decision for Cesarean birth. Oxytocin administration is certainly a valuable therapy when medical indications for induction or augmentation of labor exist. However, the dictum of *primum non nocere* must be a prime consideration in making decisions regarding its use. A portion of the increase in incidence of Cesarean birth may be related to failures of elective induction of labor. Further research is needed to evaluate this hypothesis. We believe, however, that elective induction of labor for physician convenience, or solely for presumed fetal postmaturity, has no place in contemporary obstetrics.

Dystocia (including both feto-pelvic disproportion and failure to progress in labor) has accounted for more than one third of the rise in Cesarean birth rates.[3] The goals of "undamaged infants and reasonable duration of labor" are cited by Bottoms et al.[3] as major reasons for this change. A substantial proportion of the obstetric literature[11–14] supports the contention that more liberal use of Cesarean delivery is associated with improved infant outcome. Clinical data to support this conclusion, however, are almost nonexistent. Until adequate clinical trials can show causal relationship between neonatal mortality and morbidity and route of delivery, Cesarean birth is certain to continue to come under attack. We do not intend by this discussion to underestimate or diminish the value of Cesarean birth in contemporary obstetric practice. Rather, we urge that adequately controlled trials be performed to assess and document the benefits and costs associated with it in management of specific complications of pregnancy and birth.

Electronic fetal monitoring has also been implicated as a factor in the rising incidence of Cesarean birth. Haverkamp reported a doubling of the Cesarean birth rate in monitored patients with no improvement in perinatal outcomes.[15] Banta and Thacker concluded from a review of the literature that electronic fetal monitoring (EFM) "is no better than auscultation" and that it "adds little except to increase the Cesarean section (CS) rate."[16]

Conversely, Evrard and Gold felt that "the preponderance of evidence shows that fetal monitoring is predictive of fetal hypoxia, will slightly increase the CS rate, and will decrease the perinatal mortality."[17] Until such time as adequately controlled, randomized trials have been performed to assess the value of fetal monitoring in predicting perinatal results and to compare its sensitivity and selectivity with that of less invasive techniques, the debate will continue. In our study, we found no association between the Cesarean birth rate in the PSRO area and the incidence of patients who underwent Cesarean birth previously being electronically monitored. Unfortunately, however, we did not collect data to indicate the total proportion of pregnancies in which fetal monitoring was used. This issue will also be a major focus of subsequent study.

Maternal outcomes in our study were good. Two deaths occurred in the 5,002 cases studied, a mortality rate of 40/100,000. Although this is certainly a reflection of the relative safety of the operation, it is more than three times higher than the 1976 U. S. maternal mortality statistic of 12.3/100,000. If one assumes the 40/100,000 statistic we found to be a component of the 12.3/100,000 figure for both routes of delivery, the relative mortality risk of Cesarean birth is more than five times that of vaginal delivery. This speaks to the necessity of assuring that Cesarean delivery is performed only where clearly indicated.

The other measure of maternal outcome we studied was complication rate. Using a hospital stay of greater than seven days as a proxy for complications, we found that an average of 10.2 per cent of the patients had "complications" that extended their hospitalization. However, much of the local variation in this statistic is an artifact of the well known, and unexplained, difference in average length of stay between the eastern and western parts of the United States. Thus, this finding too must be interpreted cautiously.

Infant outcomes were assessed in our study by the proxy measure of discharge with or before the mother. Although the range was large (2.2 to 15.0 per cent), the average infant complication rate (5.8 per cent) was reasonably low, considering that Cesarean birth is generally reserved for high-risk infants. The overall infant mortality in our series (15.3/1,000) is not appreciably different from the 1976 U.S. perinatal mortality ratio (17.3/1,000).

We were also pleased to note that the vast majority of women we studied (96.1 per cent) had received at least some prenatal care. In only one area (H) did as many as 20 per cent of the women who underwent Cesarean birth have no documentation of prenatal care. However, the rate of infant complications in area H (15.1 per cent) was significantly higher ($p < 0.05$, $df = 7$) than in the other eight areas. This reinforces the importance of adequate prenatal care, especially in high-risk pregnancies.

Conclusions

This study demonstrates the viability of using local PSROs as a means of conducting national studies on important issues in medical practice. Despite a number of design and sampling weaknesses that limit the strength of our conclusions, our data demonstrate that, although the rate of Cesarean births continues to rise, only an inconsequential number of these operations are inappropriate in the judgment of peer physicians in the participating hospitals. However, the relative risks and benefits of the higher rates have not been adequately examined. Further study of such concerns as the influence of oxytocin induction, economics and physician convenience, and electronic fetal monitoring on the rising rate of Cesarean birth is essential to define the proper role of this form of delivery in contemporary obstetrics. The PSRO may be a valuable vehicle for conducting such studies on a communitywide basis.

References

1. Goran MJ, Roberts JS, et al. The PSRO hospital review system. Med Care 1975;13(Supplement):1-33.
2. Jessee WF, Munier WB, et al. PSRO: an educational force for improving quality of care. N Engl J Med 1975;292:668.
3. Bottoms SF, Rosen MG, Sokol RJ. The increase in the Cesarean birth rate. N Engl J Med 1980;302:559.
4. Obstetrical practices in the United States, 1978: Hearing before the Subcommittee on Health and Scientific Research of the Committee on Human Resources, United States Senate. Washington, D.C.: Government Printing Office, 1978.
5. Pettiti D, Olson RO, William RL. Cesarean section in California—1960 through 1975. Am J Obstet Gynecol 1979;133:391.
6. Marieskind HE. An evaluation of Cesarean section in the United States. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation/Health, Department of Health, Education and Welfare, 1979.
7. Lowe JA, et al. Cesarean sections in U.S. PAS hospitals. PAS Reporter 1976; Vol. 14, December.
8. Respess J. Application of process and outcome measurements in an EMCRO. In: Evaluation in health services delivery. New York: Engineering Foundation Conferences, 1973.
9. Kappy KA, Cetrulo CL, et al. Premature rupture of the membranes: a conservative approach. Am J Obstet Gynecol 1975;134:655.
10. Dunn PM. Obstetric delivery today: for better or for worse? Lancet 1976;i:790.
11. Pakter J, Nelson F. Factors in the unprecedented decline in infant mortality in New York City. Bull NY Acad Med 1974;50:839.
12. Diddle AW, Senamer JR, Slowey JF. Cesarean section: changing philosophy. Postgrad Med 1973;53(3):160.
13. Sutherst JR, Case BD. Cesarean section and its place in the active approach to delivery. Clin Obstet Gynecol 1975;21:241.
14. Paul RH, Huey JR Jr, Yaeger CF. Clinical fetal monitoring: its effect on Cesarean section rate and prenatal mortality: five year trends. Postgrad Med 1977;61(4):160.
15. Haverkamp A, Thompson H, et al. The evaluation of continuous fetal heart rate monitoring in high-risk pregnancy. Am J Obstet Gynecol 1976;125:310.
16. Banta HD, Thacker SB. Costs and benefits of electronic fetal monitoring: a review of the literature. Hyattsville, MD: National Center for Health Services Research, 1979. (DHEW Publication No. (PHS)79-3245.)
17. Evrard JR, Gold EM. Cesarean sections: risk/benefit. Perinatal Care 1978;2(8):4.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Periodic Medical Review: The Impact in Pennsylvania

MICHAEL A. CALOGERO, M.S., NHA,* AND DANIEL B. HILL, PH.D.†

Periodic Medical Review (PMR) of all Medicaid beneficiaries in skilled nursing facilities (SNFs) in Pennsylvania was implemented in fiscal year 1973-74. In this study, the authors explored the impact of PMR through its first four annual review periods. Because over two thirds of all SNF patients who were determined (over the four-year period) to require lower levels of care were identified during the first year of the program, the study was focused on the impact of first-year reviews on patient placements and nursing home industry characteristics in the subsequent years. Although the results did not strongly show that PMR had significant long-term effects, the general conclusions reached were that: 1) first year PMR decisions altered (in the intended direction) SNF patient placements and structural characteristics of the nursing home industry in Pennsylvania; 2) the initial impact of PMR was still apparent by the fourth year, but much weaker; and 3) the PMR review team may have been more lenient in their decisions concerning appropriate levels of care for reviewed SNF patients after the first year.

DURING THE PAST DECADE, expenditures for nursing home care have been the fastest growing component of national health expenditures. From a 1965 level of about $2 billion, nursing home expenditures increased to almost $16 billion in 1978.[1] Expenditures by federal, state and local governments represented 53 per cent of nursing home expenditures in 1978[1]; federal and state expenditures under the Medicaid program were 87 per cent of all government expenditures for nursing home care.[1] Medicaid spending for nursing home care constituted almost 40 per cent of all Medicaid expenditures during 1978.[1] These data illustrate the importance of nursing home costs as a public policy issue and, in particular, the financial implications for the Medicaid program.

A number of researchers have concluded that significant portions of nursing home patients could receive appropriate care at lower levels of institutional care or in community settings if such care were available.[2-7] These findings raise concern regarding the use of Medicaid funds, but they also have implications for the inappropriately placed patient. The provision of unnecessary services may lead to physical, motivational and emotional problems for the elderly patient.[8-13]

Periodic Medical Review

In order to monitor the levels of care at which nursing home patients are placed

---

* Administrator, Olivette Skilled Nursing Care Center, Elgin, Illinois.

† Associate Professor, Department of Health Services Administration, School of Community and Allied Health, University of Alabama in Birmingham.

Address for reprints: Michael A. Calogero, M.S., NHA Administrator, Olivette Skilled Nursing Care Center, Elgin, IL 60120.

0025-7079/82/0100/0085/$01.10 © J. B. Lippincott Co.

and to determine the appropriateness of care given to Medicaid-funded patients, Congress mandated a program of Periodic Medical Review (PMR) in 1969. PMR is a state-run program involving external peer review of the care rendered Medicaid patients in skilled nursing facilities and mental hospitals.

> The primary goal of Periodic Medical Review is to ensure the provision of a range and quality of medical and nursing management, and social work support for Title XIX skilled nursing home and mental hospital patients that is both necessary and commensurate with their clinical and physical needs and optimum social functioning.[14]

Under PMR, each skilled nursing facility (SNF) that serves Medicaid patients is to have an on-site inspection by a Medical Review (MR) team at least annually. Federal regulations stipulated that the Medical Review team be composed of one or more physicians and other appropriate health and social service personnel.‡ In Pennsylvania, during the time period studied, each team consisted of one physician, one or more registered nurses and one social worker.

An inspection involves a separate review of each patient's medical records by a representative of each professional discipline on the team. Direct observation and interview (if possible) of each patient also contribute to the determination of appropriateness of placement. Group discussion of all cases results in determination of the number of and reasons for inappropriate placements. The findings of the team are presented to the administration of the facility in an exit interview, along with a description of the necessary action for correction of any deficiencies and suggestions for improvements in care and/or services to patients. Once the team decisions are presented, the patient and/or the physician have the option to appeal. If neither appeal or if the appeal is not successful, then the patient must be moved to the level of care stipulated by the MR team. This may involve a simple reclassification or a physical relocation of the patient.

Issues Concerning PMR

The potential effectiveness of Periodic Medical Review of SNF patients is affected by two major sets of issues. The first relates to the nature of the long-term care patient and his/her needs. The second set of issues involves the more general concerns raised when governmental agencies attempt to regulate the behavior of private firms.

Appropriate services for an elderly long-term care patient are much different from those for an acute-care hospital patient. An acute-care hospital is a disease-oriented facility with its resources focused on the disease rather than on the patient.[15] In contrast, the nursing home patient is usually institutionalized for an extended period of time and is dependent upon the institution for a wide range of medical, social and psychological needs.[16] Because the needs of the long-term care patient cannot be stated in strictly medical terms, it is difficult to specify standards of appropriate care that are applicable to the individual patient. There exists a substantial lack of knowledge as to what constitutes quality and/or effective care in the long-term care setting.[17,18] While a number of rating scales and patient classification schemes have been developed, definitive standards for use by external reviewers are not available.[3,10,19–21] Based on the desirability of considering the multiplicity of patient needs, the interdisciplinary team approach to medical review has received considerable support.[22,23]

As a regulatory mechanism charged with making judgments in the absence of well-defined standards, PMR has the potential to be afflicted with the problems that have been associated with other regulatory programs. A number of authors have argued that a regulatory program will tend to have a declining level of effectiveness over the life cycle of the program and may eventually benefit the regulated parties.

---

‡ A physician is no longer required as part of the on-site team, but was during the time period addressed in this study.

As Jaffee stated, "The American people have no enthusiasm for regulation. In the presence of general breakdown or glaring evils they will respond with regulation, but their concern will not survive the resolution of the crisis."[24] While the "people" may lose interest in a regulatory program, those parties most directly affected (the regulated industry) will focus their energies on trying to influence the nature and scope of regulation.[18,24,25] Faced with declining public support, sometimes evidenced by inadequate funding,[25] and cognizant of the fact that they might be blamed for any failures that occur in the regulated industry,[24,26] the regulators can become captive to the interests of the regulated industry.[24–27] This outcome, sometimes attributed to the self-preservation instincts of the regulatory agency,[24–27] is most likely to occur when the public interest that the agency is intended to serve is difficult to assess.[26]

It is not obvious whether a pessimistic prognosis for the effectiveness of Periodic Medical Review is warranted. A scenario depicting the capture of this regulatory program by the regulated industry is plausible. The program was initiated during a time of great concern over nursing home abuses and the care rendered to the elderly.[28–30] However, in contrast with many regulatory programs, PMR is charged with more than the protection of the general public interest. The program is administered by agencies (the state Medicaid agencies) that have a direct financial stake in the effectiveness of the program.

Because the study reported herein did not involve observations of the Medical Review decision-making process, the capture hypothesis was not directly tested. The authors examined the impact of PMR on patient placement patterns aggregated at the county and state levels in Pennsylvania. While the results are *consistent* with the existence of some degree of nursing home influence on the regulatory process, to interpret the findings as supportive of the capture hypothesis would be highly speculative.

Other Studies

Two prior studies have examined the recommendations of Medical Review teams under the PMR program. A Rhode Island study reported the results of Medical Review of SNF patients during the program's first year, 1973. Of the Medicaid SNF patients reviewed, 30.3 per cent were determined to be inappropriately placed. Of the inappropriate placements, 97 per cent were recommended for transfer to intermediate care facilities (ICFs), and 3 per cent were recommended for community care.[6]

During 1975, the PMR program in Massachusetts identified 18.2 per cent of all Medicaid SNF patients in the state as being inappropriately placed. The number and per cent of patients so identified varied considerably between facilities.[7]

Each of these two studies examined one dimension, review team decisions, of the PMR program during one early year of the program. While the magnitude of the inappropriate placements reported is of interest, the studies provide only limited insight concerning the effectiveness of PMR in stimulating a more efficient use of Medicaid funds. In the study reported herein, the authors attempted to capture a more complete picture of the impact of PMR in one state, Pennsylvania.

Data

The time period covered by this study corresponds to the first four annual PMR periods in Pennsylvania: fiscal years (July through June) 1973-74, 1974-75 and 1975-76, and calendar year 1977. Data on the number of SNF patients reviewed and review team recommendations were derived from summary work sheets provided by the Division of Medical Review, Pennsylvania Department of Public Welfare. Due to the Department's concern for the confidentiality of patient and facility data, totals were compiled on a county-by-county basis. No individual facility data were used. Pennsylvania has 67 counties, although only 65 had nursing homes during the time period studied. Thus, there were 65 separate totals used in this study.

Other data collected included elderly population (65 years or over) by county,[31] percentage distribution of urban and rural elderly residents by county,[32] and SNF beds and ICF beds by county.[33,34]

Hypotheses

The direction pursued in this study was based on initial observations of aggregated data for the state of Pennsylvania. As shown in Table 1, the per cent of Medicaid SNF patients who were recommended for alternative levels of care varied drastically between the initial year of PMR (1973-74) and subsequent years. The per cent of patients recommended for change in care level in the first year (34.05 per cent) exceeded the corresponding figures for the Rhode Island and Massachusetts studies.[6,7] In contrast, recommended changes accounted for no more than 6.1 per cent of patients reviewed in each of the second through the fourth years.

Based on these observations, three hypotheses were developed. Each of the hypotheses, which appear below, is consistent with the significant reduction in recommended changes in level of care following the first of PMR.

> H1. The Medical Reviews of SNF patients in 1973-74 resulted in a major relocation of SNF patients to lower levels of care and had a lasting impact on the SNFs Medicaid patient placement and retention practices in subsequent years.
>
> H2. While the Medical Reviews of SNF patients during the first year resulted in a significant relocation of SNF patients to lower levels of care, SNFs returned to their pre-PMR patient placement and retention practices after the second-year reviews.
>
> H3. PMR had no significant impact on SNF patient placement in either the short run or the long run (null hypothesis).

If H1 were valid, it could be implied that, as a result of the first year of experience with PMR, the SNFs improved their in-house review mechanisms for monitoring patient placements. If either H2 or H3 were valid, the implication is that the reduced number of recommended changes after the first year reflected modifications in the PMR criteria. Several reasons for review criteria modifications can be postulated, although it was not possible to test the validity of these explanations in this study. First, the PMR teams may have determined that first-year reviews placed too much weight on medical criteria with in-

TABLE 1. Comparative Results of Four Medical Reviews in Pennsylvania

	1973–74	1974–75	1975–76	1977
Patients reviewed (total)	18,046	16,402	19,797	23,140
Patients recommended for change in level of care (total)	6,145	869	1,225	758
% of total recommended for change in level of care	34.05	5.30	6.19	3.28

Source: Pennsylvania Department of Public Welfare, Division of Medical Review, unpublished data obtained through personal contact, 1978–79.

sufficient consideration of psychological and social factors. Second, PMR teams may have felt that the first-year reviews were not sensitive to the availability of alternative levels of care. Third, the PMR team members may have become more sensitive to the viewpoints of SNF personnel regarding patient placement. H3 carries the additional inference that the PMR program had little success in having recommended changes actually occur, either because of weak enforcement and/or as a result of successful appeals by SNFs. Data concerning compliance with PMR recommendations were not available for this study.

Another potentially important variable would be the change in review team personnel that might have occurred over the four years. Although precise information was not available, officials of the Department of Medical Review stated that there was not a significant turnover of review team personnel during this time period.

Methodology

Because 68.3 per cent of all recommended changes for Medicaid SNF patients during the first four annual reviews occurred during the first year, the per cent of Medicaid SNF patients recommended for change in level of care during the first year (CHANGE), by county, was the variable representing the focal point of this study. This variable was correlated with supply measures for the first year, and with patient population and supply measures for the second year (to determine initial impact) and the fourth year (to determine lasting impact).

The impact of PMR on nursing homes in Pennsylvania was examined by determining the rank correlations between CHANGE and two types of variables: Medicaid SNF patients and structural variables. The tau b statistic was used to measure the direction and strength of correlations.

Medicaid SNF Patients

Because all Medicaid SNF patients are reviewed annually under PMR, the number of reviews (PTSREV) during a review year provides a measure of the average Medicaid SNF patient population for the year. Use of this measure carries the implicit assumption that the reviews are evenly spaced over the year. While a number of factors might affect the number of Medicaid beneficiaries in SNFs across the state, it was felt that if PMR had a significant impact on patient placements, the *relative changes* in Medicaid SNF patients across counties during the study period should be correlated with PMR recommendations. $PTSREV_1$, $PTSREV_2$, and $PTSREV_4$ connote Medicaid SNF patients reviewed in the programs during the first, second and fourth years, respectively; $PTSREV_2/PTSREV_1$ is the ratio of second-year reviews to first-year reviews, etc.

Structural Variables

Medicaid patients occupied 40 per cent of all SNF beds in Pennsylvania during PMR's first year; this figure was 48 per cent during the program's fourth year (1977). Medicaid SNF patients recommended for lower levels of care in 1973-74 occupied 13.6 per cent of all SNF beds in the State. These figures indicate that if PMR were successful in altering patient placements, some correlation between CHANGE and subsequent growth rates for SNF beds might be anticipated. An even more pronounced (relative) impact on ICF beds could be anticipated, due to the fact that there were almost four SNF beds per ICF bed during the first year. The number of SNF patients recommended for lower levels of care in the first year, over 99 per cent of whom were recommended for ICF placement, equaled 52.7 per cent of the number of existing ICF beds.

TABLE 2. Correlations Anticipated Under Three Hypotheses; Dependent Variable: Change

Variable	Hypotheses		
	H1	H2	H3
$\text{PTSREV}_2/\text{PTSREV}_1$	−	−	N
$\text{PTSREV}_4/\text{PTSREV}_2$	N	+	N
$\text{PTSREV}_4/\text{PTSREV}_1$	−	N	N
$\text{SNFBDS}_2/\text{SNFBDS}_1$	−	−	N
$\text{SNFBDS}_4/\text{SNFBDS}_2$	$-^{a}$	+	N
$\text{SNFBDS}_4/\text{SNFBDS}_1$	−	N	N
$\text{ICFBDS}_2/\text{ICFBDS}_1$	+	+	N
$\text{ICFBDS}_4/\text{ICFBDS}_2$	$+^{a}$	−	N
$\text{ICFBDS}_4/\text{ICFBDS}_1$	+	N	N

− means negative correlation anticipated.

+ means positive correlation anticipated.

N means no significant correlation anticipated.

[a] means correlation anticipated if lag time was necessary for full impact on bed supplies.

The supplies of SNF beds and ICF beds in each county, measured on a per-100-elderly (65+) population basis, are represented by SNFBDS and ICFBDS, respectively. (In 1975, less than 3 per cent of Pennsylvania nursing home patients were under the age of 65.[35])

TABLE 3. Tau b Values Measuring Correlations of Selected Variables with Change

Variable	Tau b Value
SNF Beds/100 Elderly Population	0.16477*
ICF Beds/100 Elderly Population	−0.05682
Hospital Beds/1,000 Total Population	−0.10440
$\text{PTSREV}_2/\text{PTSREV}_1$	−0.32955†
$\text{PTSREV}_4/\text{PTSREV}_2$	0.08062
$\text{PTSREV}_4/\text{PTSREV}_1$	−0.12074
$\text{ICFBDS}_2/\text{ICFBDS}_1$	0.04372
$\text{ICFBDS}_4/\text{ICFBDS}_2$	0.08668
$\text{ICFBDS}_4/\text{ICFBDS}_1$	0.14988*
$\text{SNFBDS}_2/\text{SNFBDS}_1$	−0.04121
$\text{SNFBDS}_4/\text{SNFBDS}_2$	0.02137
$\text{SNFBDS}_4/\text{SNFBDS}_1$	−0.02060

* Statistically significant at the 0.1 level.

† Statistically significant at the 0.001 level.

Two considerations should be raised regarding the relationships between CHANGE and structural variables. First, the relationships might not be expected to be as strong as that between CHANGE and PTSREV. Some modifications in Medicaid patient placement might be accommodated through altered levels of excess capacity and changes in non-Medicaid patient census. Second, one might question the impact that PMR could have on nursing home construction during the time framework covered by this study (roughly 3½ years after the first review). However, the possibility exists that some nursing homes could alter the number of beds in existing facilities, while (more importantly) others might reclassify beds from SNF to ICF.

The results to be anticipated under each of the three hypotheses are presented in Table 2. In general, the correlation of CHANGE with the ratio of the values of a variable in the first two years is used to represent the initial impact of PMR; an opposite direction of correlation with the fourth year to second year ratio represents a reversal of the initial impact; and a fourth year to first year ratio sharing a similar correlation, as does the ratio for the first two years, represents a retention of the initial impact. The ratio of SNF Medicaid patients reviewed between two years is used as an indication of the relative impact of PMR across counties. This implies that factors other than PMR that might have affected SNF patient population trends had uniform or, at least, random effects across counties.

The tests represented in Table 3 reflect the major thrust of this study. Additional tests, conducted to gain further insight from the available data, are reported under in the "Additional Analyses" section of this article.

Results

The results derived from testing the three hypotheses are summarized in Table

3. Table 4, which will be used to complement the interpretation of Table 3, provides relevant data for the state.

The most direct impact of first-year PMR on patient placement would be reflected in SNF Medicaid censuses in the second year. In both the state data and the test using county data, the findings support the conclusion that PMR affected the placement of SNF patients. For the state, the SNF Medicaid population decreased by 9.11 per cent between 1973-74 and 1974-75 (Table 4). The fact that the SNF Medicaid population was 25.72 per cent *higher* in 1977 than during 1973-74 indicates that the second-year reduction was counter to the trend, lending support to the conclusion that the second-year figure demonstrates a response to first-year PMR decisions. The correlation of CHANGE with $PTSREV_2/PTSREV_1$, using county data, provides further support for this conclusion (Table 3). The relationship is in the anticipated direction (negative) and significant at the 0.001 level.

The findings concerning structural variables in the first two years of PMR are consistent with H1 and H2. As shown in Table 4, there was a sizable increase in ICF beds (38.74 per cent) in the second year, while SNF beds increased by a modest amount (2.88 per cent) in the second year. The tested correlations of CHANGE with the second year to first year bed ratios produced relationships in the anticipated direction, but these relationships were very weak and not statistically significant.

Comparative analyses of data for year 4 (calendar year 1977) with data for years 1 and 2 provided tests of H1 versus H2. H1 predicts a retention in year 4 of the impact of first-year reviews as witnessed in year 2; H2 predicts a dissolution of that impact. The findings do not provide a firm basis for

TABLE 4. Change in Selected Variables—Years 1, 2, and 4
Total Group—All Counties

Selected Variables	Year 1 (1973–74)	Year 2 (1974–75)	Year 4 (1977)
Patients reviewed* (total)	18,046	16,402	23,140
% change from year 1	—	− 9.11	25.72
% change from year 2	—	—	41.08
SNF Beds† (total)	45,084	46,381	48,160
% change from year 1	—	2.88	6.82
% change from year 2	—	—	3.84
ICF Beds† (total)	11,665	16,180	15,694
% change from year 1	—	38.70	34.54
% change from year 2	—	—	− 3.03

* Pennsylvania Department of Public Welfare, Division of Medical Review. Unpublished data obtained through personal contact, 1978–79.

† Pennsylvania Department of Public Welfare. Unpublished data obtained through personal correspondence, Februrary 1979.

accepting one hypothesis over the other. The correlations between CHANGE and PTSREV$_4$/PTSREV$_1$, SNFBDS$_4$/SNFBDS$_1$ and ICFBDS$_4$/ICFBDS$_1$ are all in the direction anticipated under H1, but only the ICF variable provides a relationship that is statistically significant at the 0.1 level. For the other two variables, the correlations between CHANGE and the fourth year to second year measures are in the opposite direction but weaker than the second year to first year correlations. From Table 4, it can be seen that, as in year 2, the relative growth in the number of ICF beds between years 1 and 4 was much greater than the growth in SNF beds for the state. All of the growth in the number of ICF beds occurred between years 1 and 2.

Additional Analyses

Three additional types of analyses were conducted in an attempt to shed additional light on the findings. First, tests were conducted using subsamples of counties, urban and rural. Second, the correlation between CHANGE and CHANGE$_2$, the corresponding measure of PMR decisions in the second year, was determined. Finally, correlations between CHANGE and nursing home bed supplies during the first year were examined.

Urban/Rural

The counties in Pennsylvania were classified into two groups, urban and rural. Those counties having 55 per cent or more of their elderly population defined as rural[36] were classified as rural counties. The dividing line was chosen to provide equivalent numbers of counties in each subsample.

The tests conducted for the original sample were again performed for each subsample. The purpose of this analysis was to determine whether the subsamples could reinforce the findings derived from the larger sample. A second rationale for examining the two subsamples was provided by the findings of Davis and Gibbin.[37] In a study of nursing home residents in the Buffalo, New York area, Davis and Gibbin determined that the percentage of urban patients judged to be inappropriately placed was twice that of rural patients. They attributed the more common misuse of urban facilities to the less extensive social support network (family and friends) characteristic of the urban elderly population. The subsample findings are summarized below.

1. The most direct test of the initial impact of PMR is the correlation between CHANGE and PTSREV$_2$/PTSREV$_1$. The negative correlation observed for the total sample was substantiated for the urban subsample (−0.258 tau b, 0.05 significance level) and for the rural subsample (−0.414 tau b, 0.01 significance level).
2. The test of the rank correlation between CHANGE and PTSREV$_4$/PTSREV$_1$ indicated some retention of the initial PMR impact in rural counties (−0.251 tau b, 0.1 significance level), but no retention in urban counties (0.0 tau b).
3. The finding of a positive correlation between CHANGE and ICFBDS$_4$/ICFBDS$_1$ in the total sample was attributable only to the rural counties (0.387 tau b, 0.01 significance level).
4. Both subsamples were consistent with the total sample in a lack of significant correlations between CHANGE and the SNF beds variables.

In general, tests applied to the total population and to each of the subsamples support the rejection of H3, the null hypothesis. The findings for urban counties were consistent with H2. The findings for rural counties supplied some support for H1.

The lack of a retained impact on patient placements in urban SNFs might have resulted from two factors. First, social support considerations, as noted by Davis and Gibbin, may have assumed increased importance in the placement of urban patients. Second, the finding that urban ICF bed supplies were not responsive to PMR decisions indicates a lack of institutional alternatives for the urban SNF patients.

Second-Year Recommendations

The rank correlations between recommended changes in patient placements during the first and second year of PMR were examined for two reasons. First, it was felt that the consistency of review decisions might be reflected in this relationship. Second, it was recognized that the weaker associations between CHANGE and SNF Medicaid patients in the fourth year might be attributed to a reversal of the comparative levels of recommended changes across counties after the first year.

To address directly the second point above and to provide an initial test of the first point, the relationship between CHANGE and $CHANGE_2$ was tested. Because of the considerable difference between the per cent of SNF patients recommended for relocation in the first year (34.05 per cent) and the reduction in SNF patients between the first and second year (9.11 per cent), it was felt that consistency in the review criteria would be reflected in a positive relationship between CHANGE and $CHANGE_2$. On the other hand, a negative relationship between CHANGE and $CHANGE_2$ would provide an explanation for the weaker relationship between CHANGE and the patient placement variable ($PTSREV_4/PTSREV_1$) for the fourth year. The observed correlation between CHANGE and $CHANGE_2$ (0.06300 tau b) was insignificant, providing support for neither consistency in review criteria nor a reversal of the relative impact of first-year reviews. No significant correlation was observed in either the urban county subsample or the rural county subsample.

A second test of the correlation between first-year and second-year reviews was conducted to provide further analysis of the apparent consistency of review criteria. For this test, a variable was constructed to represent the *expected* number of Medicaid SNF patients reviewed in year 2 ($EPTSREV_2$). For each county, the expected reduction in patients reviewed was computed by multiplying the number of patients recommended for change in the *county's* SNFs during the first year (CHANGEREC) by the *state* ratio of second-year reduction in SNF patients reviewed (1,644) to first-year recommended changes (6,145), i.e., $EPTSREV_2 = PTSREV_1 - (CHANGEREC) \times (1{,}644/6{,}145)$. A value of actual second-year patients reviewed in excess of $EPTSREV_2$ for a county would indicate that the SNFs in that county had altered patient placements, relative to the level of recommended changes, to a lesser extent than had SNFs across the state. Thus, if the PMR review criteria were consistent in the first two years, one would expect a positive correlation between $CHANGE_2$ and $PTSREV_2/EPTSREV_2$. On the other hand, a negative correlation would indicate a change in review criteria during the second year and would suggest that the relative impact across counties of PMR decisions in the first year might have been reversed after the second year. In testing this correlation, no relationship was found between $PTSREV_2/EPTSREV_2$ and $CHANGE_2$ (−0.00073 tau b), a finding also observed for the urban and rural subsamples.

In summary, the tests linking PMR reviews in the first two years provide no evidence that second-year reviews *directly* modified the impact of first-year reviews on SNF patient placements. However, it appears that second-year review decisions were not affected by the responsiveness of SNFs to recommendations made during the first review period. This suggests that more lenient review criteria were applied in the second year. If this were the case, it might indicate that a) PMR criteria in the first year were unrealistically narrow and applied too rigidly and/or b) the PMR program was less effective in identifying inappropriately placed SNF patients during the second year.

Change and Bed Supplies

The alternative explanations presented above (a and b) bring to light a critical consideration in judging the PMR program after its initial year. If one is to judge the effectiveness of the program in its first year by its impact on patient placements, then one must assume that the program effectively identified *inappropriately* placed patients. This assumption cannot be tested directly with available data. An indirect assessment can be made by correlating first-year decisions with variables that are considered to be determinants of inappropriate utilization. Several authors have concluded that "Reomer's law," an availability effect, exists with respect to the utilization of inpatient hospital care.[38-40] According to these researchers, the demand for inpatient care is positively correlated with the available supply of beds, and patients for whom the need for treatment in an acute care facility is questionable will represent a greater portion of patient populations in areas with a relative abundance of beds. The logic behind the availability effect can be extended to the nursing home bed supplies. However, a study by Willemain and Farber[41] provided somewhat tentative findings that contradicted the existence of an availability effect with respect to inappropriate placement of patients in SNFs. Their study used PMR decisions concerning the appropriateness of SNF placements in eight health planning regions in Massachusetts. Controlling for ICF beds and hospital beds, they observed a negative correlation between SNF beds (per 1,000 elderly) and inappropriate placements in SNFs. Second, they found an inverse correlation between utilization of SNFs and the supply of ICF beds.

In this study, the relationships between CHANGE and nursing home bed supplies in 1973-74 in Pennsylvania were tested. The bed supply variables were specified as beds per 100 elderly (65+) residents. The results did not indicate strong correlations, but the observed correlations were consistent with the existence of availability effects and contrary to the findings of Willemain and Farber. A positive correlation was observed between CHANGE and SNF beds (0.165 tau b, 0.1 significance level); a negative correlation was observed between CHANGE and ICF bed supplies (−0.057 tau b, not significant at the 0.1 level). The direction of these relationships held for the urban and rural subsamples, with the rural subsample having a stronger CHANGE–SNF relationship and the urban subsample having a stronger CHANGE–ICF relationship.

These results provide support, albeit indirect and somewhat weak, for the conclusion that SNF patients identified as being inappropriately placed by first year PMR reviews did include a more than proportionate share of all Medicaid SNF patients who were "inappropriately" placed in SNFs. It should be noted that this conclusion does not negate the possibility that first year review criteria were too rigid or applied too aggressively, e.g., that the reviewers were not sensitive to social or emotional variables.

Conclusions

Periodic Medical Review of all Medicaid beneficiaries in skilled nursing facilities in Pennsylvania was implemented in fiscal 1973-74. In this study, the authors explored the impact of PMR through its first four annual review periods. Because over two thirds of all SNF patients who were determined (over the four-year period) to require only lower levels of care were identified during the first year of the program, the study was focused on the impact of first-year reviews on patient placements and nursing home industry characteristics in the subsequent years. In assessing the conclusions listed below, the reader should bear in mind the qualifications raised previously concerning the lack of statistical significance for many of the tests. Three major conclusions are offered in descending order of the strength of the tests upon which they are based.

1. First year PMR decisions altered (in the intended directions) SNF patient placements and structural characteristics of the nursing home industry.

This conclusion is supported by aggregate data for the state and cross sectional data using county units. The first year reviews stimulated the relocation of SNF patients (and/or modifications in admissions practices) and the growth in ICF bed supplies relative to SNF bed supplies.

2. The initial impact of PMR decisions was still apparent, but weaker, during the fourth annual review period.

The declining impact of PMR after its second year raises questions concerning the long run cost effectiveness of the program.

3. PMR teams were more lenient in their appropriateness decisions concerning SNF patients after the first year of the program.

The evidence does not support the hypothesis that the drastic decline in relocation recommendations after the first year can be totally explained by reformed behavior by SNFs regarding the placement of Medicaid patients. It cannot be determined whether the greater leniency reflected the adoption of more appropriate review criteria. However, the apparent lack of association between second-year review decisions and the responsiveness of SNFs to first-year decisions lends support to the hypothesis that the effectiveness of the program was reduced after the first year.

Ultimately, the effectiveness of Periodic Medical Review rests on its success in accomplishing (or stimulating) three results:

1. The identification of inappropriately placed patients;

2. The relocation of these patients to appropriate settings; and

3. The adoption by SNFs of appropriate admission and discharge policies.

While evaluation of the PMR program by the first criterion represents an imposing research challenge, the second criterion is more amenable to empiric inquiry. The concerns raised by this study regarding the long-term effectiveness of PMR provide strong support for the advisability of tracking SNF patients who have been recommended for relocation to determine their subsequent distribution by loci of care. Knowledge concerning the placement of patients recommended for relocation during PMR's first year, the reasons behind cases where relocation did not occur and the nature of PMR decisions regarding these patients in subsequent years could greatly enhance the initial insights produced by the study reported herein.

References

1. Gibson RM. National health expenditures. Health Care Finance Review 1979;4:1.

2. Davis JJ, Gibbin MJ. An areawide examination of nursing home use, misuse, and non-use. Am J. Public Health 1971;61:1146.

3. Williams FT, Hill JG, Fairbank ME, Knox KG. Appropriate placement of the chronically ill and aged: a successful approach by evaluation. JAMA 1973;226:1332.

4. U.S. Department of Health, Education and Welfare. Problems in providing proper care to Medicaid and Medicare patients in skilled nursing facilities. Washington, D.C.: U.S. Government Printing Office, 1971. (Report No. B-164031(3).)

5. Zimmer JG. Characteristics of patients and care provided in health-related and skilled nursing facilities. Med Care 1975;12:992.

6. Allison-Cooke S, Thornberry H. Factors affecting nursing home medical review: implications for program and facility planning. Med Care 1977;15:494.

7. Connelly K, Cohen PK, Walsh DC. Periodic Medical Review: assessing the quality and appropriateness of care in skilled nursing facilities. N Engl J Med 1977;296:878.

8. Rowland KF. Environmental events predicting death for the elderly. Psychol Bull 1977;84:349.

9. Beverly EV. Nursing homes: matching the facility to the patients' needs. Geriatrics 1976;31:100.

10. Katz S, Amasa BF, et al. Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963; 185:914.

11. Schulz R. Effects of control and predictability and the physical and psychological well-being of the institutionalized aged. J Pers Soc Psychol 1976;33:563.

12. Kahana E, Coe RM. Alternatives in long-term care. In: Sherwood S, ed. Long-term care: a handbook for researchers, planners, and providers. New York: Spectrum Publications, Inc, 1975:511-72.

13. Seligman ME. Helplessness: on depression, development, and death. San Francisco: Will Freeman and Company, 1975.

14. U.S. Department of Health, Education and Welfare. Medicaid: medical review in skilled nursing homes and mental hospitals: guidelines. Washington, D.C.: Medical Services Administration, 1972.

15. Gaynes NL. A logic to long-term care. Gerontologist 1973;13:277.

16. Yokie JA. A comparison: long-term care versus acute care. Nursing Homes 1977;27:18.

17. Kosberg JI. Making institutions accountable: research and policy issues. Gerontologist 1974; 14:510.

18. Ruchlin HS. A new strategy for regulating long-term care facilities. J Health Polit Policy Law 1977;2:190.

19. Densen PM, Jones EW. The patient classification for long-term care developed by four research groups in the United States. Med Care 1976;14(supplement):126.

20. MacDonnell JA. Canadian experience with patient care classification. Med Care 1976;14(supplement):134.

21. Anderson NN, Stone LB. Nursing homes: research and public policy. Gerontologist 1969;9:214.

22. Sherwood S. Long-term care: issues, perspectives, and directions. In: Sherwood S, ed. Long-term care: a handbook for researchers, planners, and providers. New York: Spectrum Publications, Inc, 1975:3–79.

23. Kosberg KI. Methods for community surveillance of geriatric institutions. Public Health Rep 1975;90:145.

24. Jaffee LL. The effective limits of the administrative process: a reevaluation. Harvard Law Review 1954;67:1105.

25. Bernstein MJ. Regulating business by independent commission. Princeton: Princeton University Press, 1955.

26. Noll RG. The consequences of public utility regualtion of hospitals. In: Institute of Medicine. Controls on health care. Washington D.C.: National Academy of Sciences, 1975.

27. Fainsod M. Some reflections on the nature of the regulatory process. In: U.S. Senate. Separation of powers and the independent agencies: cases and selected readings. Washington, D.C.: U.S. Government Printing Office, 1970. (Document No. 91-49.)

28. Mendelson MA. Tender-loving greed: how the incredibly lucrative nursing home industry is exploiting America's old people and defrauding us all. New York: Alfred A. Knopf, Inc, 1974.

29. Townsend C. Old age: the last segregation. New York: Grossman Publishers, 1971.

30. U.S. Senate, Subcommittee on Long-Term Care. Nursing home care in the United States: failure in public policy: introductory report. Washington D.C.: U.S. Government Printing Office, 1974.

31. Pennsylvania Office of State Planning and Development. Pennsylvania population projection series: population and labor force. Harrisburg, 1973.

32. Pennsylvania Department of Public Welfare, Office for Aging. Serving older persons in Pennsylvania: a summary report and proposals for the state plan on services for older people. Harrisburg, 1977.

33. Pennsylvania Department of Commerce, Bureau of Statistics, Research, and Planning. Pennsylvania statistical abstract. Harrisburg, 1977.

34. Pennsylvania Department of Commerce, Bureau of Statistics, Research, and Planning. Pennsylvania statistical abstract. Harrisburg, 1978.

35. Pennsylvania Office of the Budget, Division of Program Planning and Evaluation. Nursing homes in Pennsylvania: a patient profile. Harrisburg, 1975.

36. Pennsylvania Department of Public Welfare, Bureau for the Aging. The elderly in Pennsylvania: data packet no. 1: demographic information. Harrisburg, 1974.

37. Davis JJ, Gibbin MJ. An areawide examination of nursing home use, misuse, and nonuse. Am J Public Health 1971;61:1146.

38. Roemer M. Bed supply and hospital utilization: a natural experiment. Hospitals 1961;35:36.

39. Feldstein MS. Economic analysis for health service efficiency: econometric studies of the British national health service. Chicago: Markham Publishing Company, 1968.

40. Anderson JG. Demographic factors affecting health service utilization: a causal model. Med Care 1973;11:104.

41. Willemain TR, Farber ME. Nursing homes and the Roemer-Feldstein hypothesis. Med Care 1976;14:880.

MEDICAL CARE
January 1982 Vol. XX, No. 1

Alcoholism, Morbidity and Care-Seeking

The Inpatient and Ambulatory Service Utilization and Associated Illness Experience of Alcoholics and Matched Controls in a Health Maintenance Organization

SANDRA L. PUTNAM, M.SC.

The issue of whether outpatient treatment of alcoholism is cost-advantageous, in the long run, for health maintenance organizations (HMOs) depends in part on whether alcoholics represent a relatively heavy burden on the health care delivery system. To clarify this issue, the paper asks whether alcoholics utilize HMO services at higher rates and in different ways than do other HMO members, and whether alcoholics experience more illnesses and injuries associated with their service utilization. The study subjects were alcoholics, identified during one year as new clients of an HMO's counseling department, which houses an alcoholism treatment component. Study subjects were matched with controls on the basis of sex, age, date of HMO enrollment, type of membership and family size. HMO medical records provided data on service utilization and associated morbidities during the three-year study period. Alcoholics were found to have utilized about 50 per cent more of all HMO services studied than did their matched controls. Alcoholics were especially high utilizers of more expensive, inpatient services. Psychosocial problems and problems classified as accidents, poisonings and violence were much more likely to underlie alcoholics' hospitalizations, and, to a lesser extent, their outpatient noncounseling utilization, than was the case for controls. Some tendency was noted for more chronic illnesses to be associated with alcoholics' service utilization, and more acute illnesses with controls' utilization.

ALCOHOLISM is frequently characterized as an illness, the economic, medical and social costs of which reach beyond the alcoholic individual and his or her social network to affect the larger community, through alcohol-related motor vehicle and other accidents, crimes, sickness absenteeism, lost earnings and lowered productivity in the work place, and high health care costs. Some researchers have tried to estimate the actual costs of alcoholism and the benefits obtained from its treatment.[1-3] However, very little has been written to document how alcoholics incur high health care costs, other than through hospitalizations and sickness and disability benefits, and how they differ from other consumers of medical care services.

The purpose of this paper is to address the cost issue indirectly by examining some of the results of a medical care utilization research project. The study was conducted at Rhode Island Group Health

---

From the Occupational Alcoholism Project, Rhode Island Group Health Association, 530 North Main Street, Providence, Rhode Island 02904.

This paper was presented in modified form at the 1980 Group Health Institute in Boston on May 6, 1980, where it was entitled, "A Comparison of the Utilization of HMO Services by Alcoholics and Matched Controls."

Supported by grant 5H84-AA-02210 from the National Institute on Alcohol Abuse and Alcoholism.

0025-7079/82/0100/0097/$01.75 © J. B. Lippincott Co.

Association, a health maintenance organization (HMO) in Providence, Rhode Island. It is a partial replication of the Genesee Valley Group Health Association's utilization project, reported by Wersinger and his colleagues.[4]

The principal research question is whether alcoholic HMO members utilize more medical care services for more and different types of illnesses and injuries than do matched controls. Such an inquiry represents one way of determining whether alcoholics place a relatively heavy burden on a closed health-care delivery system, such as an HMO. This would be expected in light of their well-documented tendency to experience high rates of certain illnesses and injuries secondary to, or associated with, alcoholism.[5] For example, among the chronic, degenerative diseases found to be positively associated with alcoholism in one non-hospital-based morbidity study were cirrhosis of the liver and hypertension, as well as peptic ulcer, asthma, diabetes, gout, neuritis, cerebrovascular disease and heart disease.[6] Other alcohol-related morbidities identified in studies using hospitalized subjects include, besides most of the above, other liver diseases, pancreatitis, tuberculosis, certain cancers, gastrointestinal problems, such as gastritis, seizure disorders, and accidents and traumas.[7–12] Many of the same illnesses are directly associated with alcoholics' mortality rates. These have been consistently found to be two or three times higher than expected rates, with excess deaths of alcoholics in most, if not all, major disease categories.[13–15]

If ambulatory, inpatient, and other HMO service utilization studied is significantly higher among alcoholic study subjects than among controls, then the need for third-party insurance coverage of alcoholism treatment and the desirability of establishing alcoholism treatment programs in HMOs may be more readily demonstrated. If differences in the nature and prevalence of morbidities associated with service utilization are found, then elucidating the physical and emotional problems of which alcoholics are at very high risk may constitute a practical means of signaling possible alcoholism among patients. The illnesses and injuries, which are viewed here as "signals," often may serve as smoke screens, obscuring the primary problem of alcoholism from medical care providers, who are already treating many alcoholics under other diagnoses.[16]

Identifying these signals of possible alcoholism may increase the chances of its early detection and timely intervention in the disease process. The chances may also be enhanced of treating the secondary or associated illnesses effectively, whether they are exacerbated by alcohol or are directly caused by the hidden, untreated alcoholism. While alcohol is clearly a dangerous drug, nevertheless, the behaviors associated with alcoholism, such as other drug use, heavy cigarette smoking, poor diet, social isolation and other lifestyle factors, may all contribute to explaining the relationship of alcoholism with excess morbidity and mortality.

Background of the Study

The HMO Setting

Rhode Island Group Health Association (RIGHA) is the only federally qualified HMO in Rhode Island. It was incorporated in 1968 and instituted services in 1971. Enrollment has increased from 1,200 in 1971 to 12,000 in 1974 to 30,000 in January 1979. The main HMO health center is staffed by 300 employees, including 27 salaried physicians. It houses, besides business and administration offices, a full laboratory and x-ray facility, a pharmacy, a dental department, vision services, and a multi-specialty area, where clinics are held in dermatology, allergy, orthopedics and other specialties. Departments of pediatrics, obstetrics-gynecology, internal medicine, surgery and counseling are also located at the health center, as is the urgent visit unit, providing 24-hour telephone coverage for emergency care, and the in-house occupational alcoholism project.

HMO coverage costs, on average, about $20 more a month than the experience-rated Blue Cross-Blue Shield plan. Most members join through employer groups, for which there are no pre-existing conditional requirements and of which there are about 900 offering RIGHA coverage. This coverage includes 100 per cent of all ambulatory care delivered by HMO providers with a few exceptions, such as most dental services and counseling services, for which fees are charged to members. These services are excluded from this study, since their inclusion would change the situation that otherwise pertains, whereby the effects of income differentials are controlled. HMO coverage also includes 100 per cent of semi-private hospitalization, together with out-of-state hospital coverage with health plan authorization. Detoxification and rehabilitation of alcoholics at area inpatient and residential facilities are covered under this benefit. Referrals by RIGHA providers to non-plan specialists are also fully covered. The extent of unauthorized, out-of-plan utilization of health care services by RIGHA members is unknown, but it is thought to be minimal.

Characteristics of the HMO over the Study Period

Over the study period, which consists basically of three calendar years—1976, 1977 and 1978— the health center was relocated to a larger and more centralized site, and a satellite clinic was opened 12 miles away. The number of non-primary physicians has increased through the addition of several part-time medical specialists, including a neurosurgeon, a urologist, and an ear, nose and throat specialist, who hold weekly in-house clinics and are paid by capitation. This trend has been accompanied by a decrease in the number of very expensive external referrals, that is, those made to non-HMO providers, usually specialists. Such external referrals have decreased by half over the study period—from 8.1 per 1000 enrollees in 1976 to 3.8 per 1000 in 1978. The number of primary physicians in ob-gyn and medicine has remained about the same, ranging from 16 to 19, over the study period, with a growing tendency to use more paraprofessionals to deliver primary medical care services. Indeed, the number of ancillary personnel—physicians' assistants, nurses, nurse practitioners and health assistants—has doubled, from 13 in 1976 to 25 in 1978. Finally, ambulatory utilization during the study period increased from about 3.2 visits per person per year in 1976 to 4.3 visits per person per year in 1978, with hospitalization rates decreasing from 402 hospital bed days per 1000 members in 1976 to 355 bed days per 1000 in 1978.

The Occupational Alcoholism Project

Fairly unique additions to the services available in this HMO setting are those provided through the outreach and treatment components of the occupational alcoholism project. The project has been funded by a grant from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) since September 1975. It has received accreditation from the Joint Commission on the Accreditation of Hospitals since 1977. It uses the employee assistance approach to the identification, confrontation and referral, by supervisors, of employees with job performance problems, which frequently mask alcohol problems. The outreach component of the occupational alcoholism project is the subject of an earlier research report.[17]

The project uses the HMO's counseling department as the principal treatment resource. Complete intake, evaluation and outpatient treatment services are offered by 12 therapists to HMO members and to employees of organizations with employee assistance programs for alcohol as well as for psychiatric problems. The alcoholism treatment program, then, is integrated into the mental health department of the HMO. It is associated with the outreach component of the occupational alcoholism project. It operates within the closed health care delivery system of the HMO, insuring information-sharing through medical records and easy access to medical care providers, with a family-oriented approach. All of these conditions are thought to increase the likelihood of early detection and intervention in alcoholism in this setting.

The counseling department has grown from serving 656 new clients or "intakes" in 1976 to 714 in 1978. An average of 63 new clients a month came to counseling during the study period. Thirty per cent of these intakes, on average, were diagnosed as having the primary problem of alcoholism. Another 27 per cent of intakes were spouses or other family members of alcoholics. Such persons are classified as having "alcohol-related" problems. Of the alcohol-involved intakes (including alcoholics and family members), about one quarter were referred directly by supervisors at their place of employment. One third were referred through a health care provider, and 37 per cent were self-referred or were referred by family or friends.

The presence of the occupational alcoholism project in the HMO made it relatively easy to identify a group of alcoholics, who were HMO members and for whom, therefore, complete medical and counseling records were available. Wersinger and his colleagues[4] lacked such ease of access to an alcoholic sample. In addition, the diversity of referral sources of alcoholics to the counseling department may insure that less severely troubled alcoholics were included in the study group than are often caught in samples of alcoholics. One possible indication that relatively early detection of alcohol problems is occurring at RIGHA is the fact that, on average, study subjects are younger (mean age is 37.3 years) than either the inpatient alcoholics (mean age is 45.8 years) or the confirmed alcoholics (with a mean age of 41.1 years) who served as study subjects in the Genesee Valley Group Health Association project.[4] Table 1 shows the age and sex distribution of alcoholics in the present study.

Relevant Literature

Studies of Alcoholics' HMO Service Utilization

Whether identified alcoholics utilize prepaid medical care services at higher rates and in different ways than do other HMO members has been the topic of two previous studies in the United States. The first, based on Boyajy's published reports,[18, 19] concerns a four-year research and demonstration project undertaken by the Group Health Association of America and funded by a grant from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Four geographically distant health plans provided data on a variety of subjects, including the utilization behavior of identified alcoholics. Although the failure to use matched control groups flawed the research, as the author notes, still, the findings were as expected.[19] The alcoholics consumed services at a significantly higher rate than other group members (ref. 19, page 419). The overall utilization rate of alcoholics at one plan, for instance, was 22 per cent higher than that of the general health plan population, controlling for age. Further, the patterns of use were found to be different for the alcoholics, who showed a predominance of unscheduled and other relatively expensive visits, including high utilization of physician services and emergency visits (ref. 18, pages iii–iv).

Boyajy also hypothesizes that alcoholics will tend to have more illnesses with a "high emotional component" than other HMO members. In support of this contention, he lists the ten most frequent diagnoses associated with alcoholics' utilization of services. One of these is anxiety, and

TABLE 1. Comparison of the Age–Sex Distributions of Study (and Control) Subjects with that of the Total RIGHA Membership during 1979

Age Groups	Age–Sex Distributions								
	Study (Control) Group*						Overall RIGHA Membership†		
	Males		Females		Total		Males	Females	Total
	#	%	#	%	#	%	%	%	%
15–24	4	5.3	5	16.7	9	8.5	25.0	27.6	26.3
25–44	51	67.1	20	66.6	71	67.0	43.7	42.4	43.1
45–64	20	26.3	5	16.7	25	23.6	27.0	26.6	26.8
65+	1	1.3	0	0	1	0.9	4.3	3.4	3.8
Total	76	100.0	30	100.0	106	100.0	100.0	100.0	100.0
	$\bar{X} = 38.8$		$\bar{X} = 33.5$		$\bar{X} = 37.3$		Median = 27.8	Median = 27.2	Median = 27.6

* Because controls were matched with alcoholics by age and sex, among other characteristics, the age–sex distributions for the two groups are identical.

† These percentages of RIGHA members in age–sex groups are based on the averages of figures for two points in the study period, May 1, 1976 and December 31, 1978.

some others are illnesses that are thought to be stress-related. In descending order of frequency, these diagnoses are alcoholism, hypertension, upper respiratory tract infection, influenza, trauma, alcoholic gastritis, gout, anxiety reaction, contusion, cirrhosis, and allergic reactions. Ulcers and drug reactions join the list in the six-, twelve- and eighteen-month follow-up data.[18] While these results are neither discussed in detail by Boyajy nor compared with a similar list of morbidities for a nonalcoholic or control group, they provide a basis for comparison with this study's results. They also give rise to one of the hypotheses tested here.

Related to Boyajy's study is the Genesee Valley Group Health Association's utilization project alluded to earlier.[4] In this project, three study groups of alcoholics were drawn from the HMO population, with some difficulty, since the HMO lacked both a psychiatric or counseling department and an alcoholism treatment component. The subjects consisted of a group of 10 inpatient alcoholics, who were identified through hospital claims forms; 59 confirmed alcoholics, who were diagnosed as such by HMO providers; and 98 suspected or high-risk alcoholics, who were identified by answers to questionnaires. The subjects were matched on a number of sociodemographic dimensions to individuals in three control groups. Utilization data were provided by hospital claims forms and ambulatory encounter forms.

The authors found, as hypothesized, that the alcoholics (inpatients and confirmed alcoholics grouped together) experienced higher rates of utilization of both inpatient and ambulatory services compared with controls. The largest differences in utilization were found for inpatient hospitalizations. This is not unexpected, since the inpatient group was selected in such a way that a bias in favor of higher rates was ineluctably introduced. Still, confirmed alcoholics utilized eight times as many inpatient services as their controls. Most of the differences between alcoholics and controls were attributed to psychiatric admissions, which were 100 times higher for alcoholics than for controls, and, to a lesser extent, to medical admissions, which were four times higher for alcoholics. No differences in the rates were observed for surgical admissions.[4]

As for ambulatory utilization, measured in terms of average number of encounters per person-year over a four-year period, the alcoholics had 23 per cent higher rates than the control groups. Much of the difference in the encounter rates was due to a 50 per cent higher illness rate for alcoholic study subjects.[4] The fact that medical hospital admissions were relatively high for alcoholics and that they had higher ambulatory encounter rates for the treatment of medical problems led Wersinger and his co-authors to suggest that alcoholics are higher utilizers of services, because they are more likely to be ill than controls.

This explanation of alcoholics' relatively high utilization is reinforced by the general finding for nonalcoholic groups that need for care or poor health is the strongest predictor of physician utilization in the fee-for-service sector.[20] More specifically, it is suggested by Boyajy's finding of a relationship among degree of alcohol involvement, level of morbidity, and utilization rates. He found that the most severely troubled alcoholics, whose consumption levels were highest and whose duration of heavy drinking was longest, seemed to be higher utilizers of services and were at the greatest risk for a number of morbidities.[18]

In addition to their greater need for care, another explanation is possible for alcoholics' higher utilization of services, when age, sex, and other sociodemographic factors are controlled. It is that they report, rather than necessarily experience, more illnesses, that they are more likely than controls to seek treatment and to be compliant in following up on it, and that they, therefore, enter the reporting system more often. This notion is difficult to test, since it is founded on the concept of illness behavior, rather than illness itself, as the independent variable for explaining utilization. However, because alcoholism is said to be an illness, the denial of which is a key symptom, and because alcoholics are ". . . notoriously non-compliant,"[12] they might be more apt to avoid medical treatment than to seek it out in circumstances that could lead to their identification as chemically dependent by medical care providers. This issue is important and requires further study.

Hypotheses

Based on the literature, the hypotheses explored in this study are:

1. The alcoholic study subjects will be found to be higher utilizers than controls of all types of services studied, particularly of inpatient hospital services, even after eliminating those for alcoholism detoxification and rehabilitation, and
2. that the study subjects will be found to experience (or report) more illnesses, both physical and emotional, and more injuries than controls.

The following three criteria are used to specify and operationalize the second hypothesis:

a. that alcoholic study subjects will report more problems or morbidities than controls;

b. that alcoholic study subjects will have more problems with a high emotional component, including primarily what were coded in this study as psychosocial problems; and

c. that study subjects will tend to report more serious or chronic problems than controls.

Data and Methods

Sample Selection

The experimental group for this study was drawn from the population of new counseling clients reported as primary case members with the primary problem of alcohol on NIAAA Initial Contact Forms, between January 1, 1977 and March 31, 1978. These forms are filled out by therapists for all counseling clients on intake or first counseling encounter, both to satisfy the NIAAA grant requirements and for internal recordkeeping purposes. A further criterion for eligibility in the study group was continued membership in the HMO, with no more than a three-month lapse, during the entire 39-month study period, from January 1, 1976 to March 31, 1979. This insured that all study subjects possessed medical records, which served as the principal data source, and that utilization data were available on every subject for the entire study period and for at least one year before and one year after alcoholics' counseling intake date or its equivalent for controls.

A total of 294 alcoholics were identified when entering counseling for the first time during the 15-month pivotal period—a total of 242 during calendar 1977 and another 52 during the first three months of 1978. Of these potential study subjects, about half were eliminated, because they lacked appropriate insurance coverage, having either no insurance or other than HMO coverage. About one third (51) of the 157 RIGHA members were ineligible for inclusion, because their HMO coverage was not sufficiently extensive. The remaining two thirds of the alcoholics who were HMO members made up the study group. Parenthetically, the spouses of the study subjects and of the controls were also included in the study. The same data were collected on their service utilization as was on their spouses' utilization. For purposes of this paper, however, data on the 76 study spouses and 76 control spouses are omitted.

Controls were selected from lists of current RIGHA members, as of March 31, 1979. They were matched to study subjects on basically the same dimensions used for matching in the Genesee Valley study.[4] These are sex, date of birth (age), date of effective HMO membership, type of HMO membership (individual or family, which often indirectly indicates marital status for adults), and number of dependent children. The task of matching, although done by hand rather than by computer, was facilitated by the fact that RIGHA identification numbers are assigned, in part, on the basis of date of enrollment, and microfiche membership files are organized by identification number. Originally, several controls were matched to every study subject. The controls who were kept in the study, after matching on sex and type of membership, were matched, as closely as possible, with alcoholic study subjects by date of birth, then by date of effective HMO membership, and finally, as the least critical matching variable, on number of dependent children. One hundred and six study subjects and the same number of matched controls were included in the study.

A possible source of bias in this study is that some potential controls, as well as some potential study subjects, were HMO members who were ineligible for inclusion in the study, because they joined RIGHA after January 1, 1976 or cancelled before March 31, 1979. Cancellations are due to job changes, with consequent loss of HMO group coverage; to geographical mobility; to divorce; in rare cases, to death; and to voluntary changes of coverage due to dissatisfaction with the health plan. Alcoholics may be more likely to experience loss of HMO coverage than other plan members. Eliminating HMO "drop-outs" from the study introduces a bias in favor of stability of life circumstances and satisfaction with the plan. In the case of alcoholics, it may favor the inclusion of those who are less severely troubled or less severely impaired. Since this would lead to greater similarity between alcoholics and controls, less difference in their utilization would be expected. Therefore, any substantial differences that are observed in utilization of HMO services, measured by encounters, and in associated problems or morbidities between experimental and control groups will gain added weight.

Characteristics of Study Subjects and Controls

The age distribution of the 106 study and matched control subjects is compared with that of the overall HMO population during the study period in Table 1. It indicates that the sample underrepresents young (15–24) and elderly (65+) RIGHA members, while subjects in the early working ages (25–44) are overrepresented, compared with the plan membership as a whole. The sex distribution is different from that of the HMO overall, where females are about equally represented. For the study and control groups, 29 per cent are female. In the sample, female subjects are slightly younger than males, with a mean age of 33.5 years, compared with a mean age of 38.8 years for males. The ratio of members to subscribers is also higher in the sample than in the plan—3.2 compared with 2.3. A subscriber is an HMO member who holds the membership contract; members include all subscribers plus the eligible dependents of subscribers. Subjects have an average of 2.1 dependents and 1.4 dependent children per subscriber. The sample is thus skewed toward family memberships (married persons). Seventy per cent of subjects had family memberships and 30 per cent had individual memberships, compared with 56 per cent family memberships and 44 per cent individual memberships (excluding Plan 65+ members) for the total RIGHA population, as of January 31, 1979. The fact, mentioned earlier, that the sample alcoholics are younger, on the average, than many other groups of alcoholics who have been studied, might be reflected in comparatively low utilization rates for both alcoholics and controls, since age is associated with illness.[21,22]

Besides the group's relative "youth" and the high representation of women (one third), compared with other samples with less than 10 per cent women, another fairly unique aspect of the study group is that it is a non-hospitalized sample. This is unusual in alcoholism morbidity studies.[10] Secondly, that it includes both acceptors of HMO alcoholism treatment and non-acceptors, those who dropped out of treatment, is also exceptional. Other studies have used, as subjects, voluntary participants in long-term alcoholism treatment, who differ in unknown ways from untreated alcoholics.[8] This study group may thus represent an infrequently sampled population of alcoholics.

Measures of Utilization

The "encounter" is the main unit of analysis according to the coding system developed for use with the medical records. An encounter is defined for a subject as a face-to-face contact with an HMO provider (physician, physician's assistant, nurse practitioner, etc.) or with a non-plan or external provider. Encounters include visits for laboratory tests and x-rays, and hospitalizations. Ambulatory encounters encompass health maintenance unit visits for initial and annual physical examinations, as well as most primary care and specialty visits. Dentistry and optometry encounters were excluded, as were prenatal care clinic visits and periodic visits with nurses for the sole purpose of blood pressure checks. Reports of some such encounters are kept separately from the HMO medical records. In other cases, the reporting of such encounters in the medical records was found to be inconsistent and incomplete. Physical therapy encounters were coded when external referrals were made. HMO-based physical therapy encounters were coded if they were reported in the medical records, which is thought to be fairly often. The extent of uneven coverage of this category is unknown. No well-child and pediatrics services were used by the subjects. Counseling encounters were coded, using counseling department records rather than medical records, but they were omitted from this analysis, which is confined to medical care utilization behavior.

Due to the definition of encounters chosen, telephone calls and letters were not included in the coding, unlike the situation in Boyajy's study. Hospital visits by doctors and home visits by nurses, which occur very infrequently, were not coded, because they are not routinely reported in the medical records. All other ambulatory encounters, all external referrals, all laboratory and x-ray procedures, and all hospitalizations were coded for all subjects. Each separate encounter was coded by type or department—medical, surgical, orthopedic, etc.—by date, by associated morbidities (called "problems"), by procedures reported done, and by the type of provider seen.*

Both numbers and proportions of encounters and procedures are used as measures of utilization in this study. "Procedures" are defined to include preventive procedures (such as certain physical examinations), follow-up encounters after initial appointments, and diagnostic procedures, such as treadmill stress tests, EKGs, nerve conduction studies, etc. Also included, particularly in the inpatient utilization category, are surgical procedures. These were coded using a somewhat revised and condensed version of the Blue Shield Schedule of Relative Values.[23] Therapeutic procedures, particularly certain drug therapies, were also coded.

The overall ratio of procedures to encounters for alcoholics and controls is 0.93, as shown in Table 2. Procedures are thus fairly good proxy measures for encounters, especially for alcoholics whose overall ratio is 0.96. Using procedures to measure utilization overestimates external referrals by 8 per cent, and somewhat underestimates outpatient encounters, especially for controls. The greatest discrepancy between the two utilization measures is in the inpatient category, where, overall, 26 per cent more procedures were coded for the average encounter, with controls' utilization being more distorted by measuring procedures than alcoholics' utilization. Encounters are therefore the measure of choice here.

"Problems" are defined as including symptoms reported by the patient when the provider failed to or was unable to diagnose the illness presented, as well as providers' diagnoses and assessments. Problem codes are based on the three-digit series of codes in the eighth edition of the International Classification of Diseases (ICD).[24] The only major revisions of these codes were the substitution of other psychosocial codes for mental disorder codes and some coding changes in the category of "Symptoms and Ill-Defined Conditions."

TABLE 2. The Ratios of Procedures to Encounters by the Three Main Service Utilization Categories for Alcoholics and Controls

Service Utilization Categories	Ratios of Procedures to Encounters		
	Alcoholics	Controls	Overall
External referrals	1.08	1.08	1.08
Inpatient encounters	1.22	1.36	1.26
Outpatient non-counseling encounters	0.94	0.85	0.90
Total	0.96	0.88	0.93

---

* The way in which information was coded on hospitalizations for any encounter or admission is illustrated by the case of a hypothetical encounter, which was a medical emergency, on 8/20/77, for the problems of facial lacerations, cerebral concussion and adverse reaction to an analgesic drug. Procedures included sutures and tomography. The length of stay was 5 days. An external referral was coded, for example, as occurring at a vascular laboratory, on 11/15/78, for the problems of lower leg trauma and thrombophlebitis, the procedure being venous evaluation. Information collected for coding ambulatory encounters included the fact that the patient, for example, on 5/4/77, was seen in the HMO surgery department for a dog bite on the ankle, which required sutures and a tetanus shot, which were provided by a physician. Another example is on 1/30/76, a patient was seen in internal medicine for sleep problems, epigastric pain, and anxiety, and was given a physical examination and a prescription for flurazepam by a physician (Certain selected drugs, particularly CNS depressants, were coded in this study, when they were reported as prescribed.)

Since this is a descriptive study, group differences for the three main service utilization categories—external referrals, hospitalizations, and ambulatory encounters—are examined, using percentage distributions and indexes of dissimilarity (ID) to summarize distribution differences.† Relative risk, defined as the ratio of the alcoholic group's score or rate to that of the control group,[4] is also employed to portray group differences. For the most important comparisons, Student's t-tests for matched or dependent samples are calculated as tests of significance.[25]‡

Results and Discussion

Utilization of HMO Services by Alcoholics

The first hypothesis of this study was supported, as shown in Table 3. Overall, the alcoholic study subjects utilized one-and-a-half times more services, measured by encounters, than did controls. The magnitude of the difference between alcoholics and controls (45 per cent) is greater than was found by Boyajy (22 per cent).

As expected, the difference in service utilization was greatest for hospitalizations. Alcoholics had 143 per cent more inpatient encounters (separate hospitalization episodes) than did controls. This difference is significant at the 0.001 level (t = 3.82, df = 105). Eliminating hospitalizations for detoxification from inpatient encounters reduces the difference between the two groups only slightly, with study subjects still utilizing about twice as many inpatient services (the relative risk is 2.02). The differences remain significant (t = 4.29, df = 105).

According to the relative risks in Table 3, alcoholics made 43 per cent more ambulatory non-counseling encounters, compared with controls. This difference is significant at about the 0.001 level (t = 3.42; df = 105). Study subjects utilized 45 per cent more laboratory and x-ray services§ and 28 per cent more external referrals. External referrals comprise the utilization category that least differentiates alcoholics and controls. The group differences here do not reach significance (t = 0.933; df = 105).

Not only did alcoholics use more medical services after seeking care at the HMO, more alcoholics than controls sought such medical care in the first place. More alcoholics were users of HMO services. More controls failed to use such services. This is shown in Table 4, where not numbers of encounters, but numbers of subjects making encounters, are compared. While alcoholic study subjects and controls were similar in their nearly universal utilization of outpatient services and laboratory and x-ray procedures, about half as many controls as alcoholics were hospitalized (52 alcoholics compared with 29 controls) and about one third as many controls were referred to non-plan providers (36 per cent of controls, compared with 53 per cent of alcoholics).

Table 4 also gives the encounters per user in each utilization category over the three-year study period. Alcoholic users had more encounters, on average, than controls in three of the four service utiliza-

---

† The index of dissimilarity is interpreted as the percentage of cases that would hypothetically need to be redistributed to make two percentage distributions equal to one another. The formula is $\frac{1}{2}\sum|P_1 - P_2|$, where P = percentage distribution.

‡ For these significance tests, each pair of individuals—each alcoholic and his/her matched control—is treated as a single case. Direct pair-by-pair comparisons of encounters are made by obtaining a difference score for each pair and testing the null hypothesis that the two populations do not differ and therefore that alcoholism has no effect on utilization and morbidity. A single-sample test of the hypothesis that the standard deviation of the differences equals zero is used. T-tests are calculated, and the null hypothesis is rejected at the .05 level of significance or below.[25]

§ When a subject was referred to the HMO laboratory, a number of tests were usually ordered done at once, so that a count of procedures rather than encounters, gives a more accurate picture of the utilization of such services.

TABLE 3. Numbers of Encounters Made by Alcoholics and Controls by Service Utilization Category, Including Laboratory Tests and X-Rays, External Referrals, Hospitalizations and Ambulatory Encounters

Service Utilization Categories	Numbers of Encounters		
	Alcoholics	Controls	Relative Risk
Laboratory tests and x-rays[*]	1780[†]	1230[†]	1.45
External referrals	143	112	1.28
Inpatient encounters (hospitalizations)	107(89)[‡]	44(44)[‡]	2.43(2.02)[‡]
Ambulatory non-counseling encounters[§]	1695	1184	1.43
Total	3725	2570	1.45

[*] Laboratory tests and x-rays are counted by procedures, rather than by encounters, because a number of such tests were usually done during one encounter. Procedures are therefore a more accurate measure of utilization in this service utilization category than encounters.

[†] These laboratory tests and x-ray procedures exclude a few codes used for drugs administered during these procedures, once for an alcoholic and six times for four controls.

[‡] The figures in parentheses are the results, when alcoholism detoxification and rehabilitation hospitalizations are excluded.

[§] Ninety-one ambulatory counseling encounters made by alcoholics were excluded from the original data, as they had been mistakenly included. The results shown omit these encounters. Thirteen such encounters for controls were similarly omitted. Their inclusion raised the relative risk slightly to about 49 per cent for ambulatory encounters. It also changed the ratios of outpatient procedures to outpatient encounters for alcoholics from 0.94 to 0.95, while the ratio for controls remained unchanged at 0.85.

tion categories. These included an average of 2 hospitalizations, compared with 1.5 for control users; 16 outpatient encounters, compared with 11.5 for controls; and 17 laboratory and x-ray procedures, relative to 13 for controls. Unexpectedly, however, more external referrals were made for and/or accepted by control users than was the case for alcoholics. An average of 3 external referrals was recorded for control users, while 2.5 was the alcoholics' average. It may be that alcoholics, who are sometimes regarded as "bad," i.e., difficult and undesirable patients, are not being as readily referred to non-plan specialists by HMO providers, despite their higher morbidity. Alternatively, it may be that as many or more such referrals are made, but they are not as readily accepted by alcoholics as by controls. It is also possible that although more alcoholics were referred to external providers than controls, the latter users,

TABLE 4. Numbers of Subjects Utilizing Services (Users) and Average Encounters Per User by Service Utilization Categories for Alcoholic Study Subjects and Matched Controls

Service Utilization Categories	Alcoholics			Controls		
	No. Users	% Potential Users (106)	Encounters Per User	No. Users	% Potential Users (106)	Encounters Per User
Laboratory and x-ray procedures	105	99.1	17	95	89.6	13
External referrals	56	52.8	2.5	38	35.8	3
Inpatient encounters	52	49.1	2	29	27.4	1.5
Outpatient non-counseling encounters	105	99.1	16	103	97.2	11.5

TABLE 5. Numbers and Percentages of "Problems" or Morbidities Reported During Encounters, by Service Utilization Categories, for Alcoholic Study Subjects and Controls

Utilization Categories	Total Occurrences of Problems or Morbidities*				Relative Risk
	Alcoholics		Controls		
	No.	%	No.	%	
External referrals	200	6.6	180	9.5	1.11
Inpatient encounters	258	8.6	88	4.7	2.93
Outpatient non-counseling encounters	2558	84.8	1622	85.8	1.58
Total reported problems associated with service utilization	3016	100.0	1890	100.0	1.60

* Total occurrences of problems are counts of the numbers of separate problem codes, summed over all problem categories, for all subjects and for distinct utilization categories. Therefore, for example, alcoholics' 2558 outpatient problems included 84 occurrences of diabetes, 64 occurrences of obesity, 129 occurrences of hypertension, etc.

while fewer in number, may have been more severely ill or injured, requiring more external services. This issue deserves further study.

Parenthetically, the types of procedures that were most frequently utilized differed both by utilization category and by group. The most striking difference between alcoholics and controls was in the marked prevalence of sterilization procedures incurred by the study group. Four of the 30 alcoholic women in the study sample were hospitalized for tubal ligations during the three-year study period, compared with no controls. Similarly, twice as many vasectomies (6) were performed, on an outpatient basis, on alcoholic men as on controls (3). Because this is essentially a pilot study, no conclusions can be drawn here. However, the questions of whether alcoholic women have higher encounter rates for obstetrical and gynecological problems and whether both male and female alcoholics have more sex-related problems than their respective controls are worthy of further study.

Illnesses and Injuries Associated with Alcoholism

The second hypothesis, that alcoholics will be more likely to experience (or report) illness than controls, was supported, although more intensive data analysis is needed to further specify the relationship between levels and types of morbidity and utilization behavior.

As Table 5 indicates, alcoholic study subjects reported 60 per cent more problems associated with their overall service utilization than did controls. The largest difference between the two groups was in hospitalizations, where nearly three times as many problems were recorded for alcoholics as for controls. For each inpatient encounter, 2.4 problems, on average, were recorded for alcoholics, compared with an average of 2 problems per encounter for controls. This suggests that when study subjects were hospitalized, they were diagnosed as having more associated problems than were hospitalized controls. For

---

A number of the results reported here are slightly different from those reported in an earlier version of this paper, entitled "A Comparison of the Utilization of HMO Services by Alcoholics and Matched Controls" by Sandra L. Putnam, presented at the 1980 Group Health Institute, May 6, 1980, in Boston, Massachusetts. This is because the original results contained data on some ambulatory counseling encounters, which somewhat inflated the psychosocial problem counts. All of the results for ambulatory encounters have been revised to exclude these few counseling encounters, and all related t-tests have been recalculated.

TABLE 6. The Percentage Distribution of Problems Reported During Ambulatory Non-Counseling Encounters Classified by Bodily Systems and Other ICD Categories* for Alcoholics and Controls

Problem Categories	Percentage Distribution of Outpatient Problems		
	Alcoholic Study Subjects	Controls	Relative Risks
Psychosocial problems	33.0(26.5)†	12.4(12.2)†	6.3(4.1)†
Cardiovascular problems	12.8	11.5	2.7
Respiratory problems	12.0	21.5	1.3
Symptoms and ill-defined conditions	11.8	7.0	4.0
Accidents, poisonings and violence	7.0	13.9	1.2
Musculoskeletal problems	7.0	14.1	1.2
Digestive problems	5.9	3.9	3.6
Endocrine problems‡	2.9	1.0	6.6
Skin problems	2.7	8.6	0.75
Urogenital problems	1.8	2.4	1.8
Sensory or nervous problems	1.1	1.6	1.7
Infectious diseases	1.1	1.7	1.6
Neoplasms	0.9	0.4	4.8
Congenital anomalies	—	—	—
Total	100.0	100.0	100.0
Weighted frequency scores§	28,928(22,873)†	12,200(12,180)†	2.4(1.9)†
Index of Dissimilarity: 31.1			

* The major disease categories in ICD-8 (The Eighth Revision of the International Classification of Diseases) are used.[24]

† Figures in parentheses are adjusted percentages, excluding the code for alcoholism (code #303) from among the psychosocial codes.

‡ Obesity, problem code 277, is classified as a psychosocial problem in this study.

§ Overall weighted frequency scores are obtained by summing the total occurrences of each problem code weighted by the number of cases with one or more occurrences of the code for all codes in each ICD category. Thus, for example, 24 alcoholics had one or more reports of the problem of hypertension during ambulatory encounters. The problem of hypertension was reported a total of 129 times for these 24 alcoholics. The weighted frequency score for the problem code, hypertension, is therefore 3096.

other utilization categories, both groups reported about the same ratio of problems to encounters (1.5 to 1 for alcoholics; 1.4 to 1 for controls).

Prevalence of Psychosocial Problems among Alcoholics

Table 6 shows the distribution of problems, classified by standard ICD categories,[24] that were associated with outpatient non-counseling encounters for the sample. Based on this table, the alcoholic study subjects were particularly prone to be seen by non-counseling providers in ambulatory encounters for "psychosocial" problems. These included anxiety, depression, family and marital problems, etc. Fully one third of problems reported by alcoholics for outpatient encounters were classified as "psychosocial," compared with 12 per cent of controls' problems. This difference is highly significant ($t = 5.39$; df $= 105$; $p = 0.0001$). Cardiovascular problems, respiratory problems and ill-defined conditions represent about 12 per cent each of alcoholics' outpatient problems. Study subjects were at nearly three times higher risk of cardiovascular problems and 30 per cent greater risk of respiratory problems. They were four times more likely to report ill-defined symptoms, including nervousness and fatigue, chest pains of unknown origin, etc. Alcoholics were about 20 per cent more likely to use outpatient services for accidents and related problems and for musculoskeletal problems than controls. The former were nearly four times more likely to report or be diagnosed with digestive problems and about seven times more likely to report endocrine problems. Sensory or nervous problems and urogenital problems were about 80 per cent more common for alcoholics, while neoplasms were five times more common among them. Finally, skin problems were 25 per cent more likely to be reported for controls. Overall, the alcoholics' relative risk of reporting outpatient problems in these categories is 2.4, and the index of dissimilarity is 31 per cent.

Three psychosocial problems are among the most frequently reported for alcoholic study subjects in the ambulatory utilization category; they are alcoholism, depression and anxiety. Nervousness/fatigue, although classified as an ill-defined condition, may be construed as an emotional problem. While this problem is among both alcoholics' and controls' most commonly reported outpatient problems, alcoholics were six times more likely to report it. Alcoholism, the most common psychosocial problem for alcoholics, was recorded at least once during outpatient encounters for 36 study subjects (about 40 per cent of the study group) and 4 controls (about 4 per cent of the control group).¶ Depression was reported 12 times more often among alcoholics, while anxiety was four times more frequent. Less difference was found for obesity, which was classified among psychosocial problems. Obesity, although twice as common among alcoholics, was the most common psychosocial problem for controls, affecting 19 of them. Other psychosocial problems showed even greater relative risks for alcoholics. Sleep problems and family problems were 15 times more likely among alcoholics. Marital problems were 12 times more frequent. Job and financial problems and other general psychiatric or emotional problems occurred 6 times more often among alcoholic study subjects. Parenthetically, alcoholics (seven) were found to be 100 times more likely to be diagnosed as having physical disorders of presumably psychogenic origin than controls (one). These findings support Boyajy's contention that alcoholics tend to utilize more services for illnesses with a "high emotional component"[19] than do controls.

The same trends are apparent for problems associated with inpatient hospitalizations. Table 7 shows that more than half of all alcoholics' problems in this utilization category were psychosocial, compared with 1 per cent of controls' problems. This difference is statistically significant at the 0.001 level ($t = 3.44$; $df = 105$). This result is consistent with the Genesee Valley study's finding of a very high risk of psychiatric hospital admissions for alcoholics compared with controls. According to Table 12, which shows the most frequent inpatient problems for alcoholics, alcoholism is the single most frequent such problem requiring hospitalization. Two other psychosocial problems, depression and suicidal ideation, are also represented among the most frequently reported codes for study subjects. Controls had *no* psychosocial problems among those most frequently reported or diagnosed.*

Among alcoholics' inpatient problems, accidents and related traumas constituted 20 per cent. They were four times more

---

¶ If HMO providers identify the problem of alcoholism among controls in the same proportions as alcoholics are identified by providers' entries in the medical records (40 per cent), then as many as 11 controls might be alcoholics. This does not invalidate the control group, since controls are not defined as nonalcoholics, but as HMO members who have not been identified by the counseling department as alcoholics during the pivotal one-year-and-three-month period.

* The most common psychosocial problem associated with alcoholics' hospitalizations was alcoholism, which was reported in conjunction with the inpatient encounters of 12 study subjects (WFS = 336). Depression was associated with 5 study subjects' hospitalizations (WFS = 40); suicidal ideation with 3 (WFS = 24); obesity with 4 (WFS = 16); and marital problems with 3 (WFS = 12). The only two psychosocial problems that were reported in conjunction with controls' inpatient encounters were obesity for one control and anxiety for one control.

TABLE 7. The Percentage Distribution of Problems Associated with Inpatient Service Utilization for Alcoholics and Controls, Classified by Bodily Systems Involved and by Other ICD Categories*

Problem Categories	Percentage Distribution of Inpatient Problems		
	Alcoholic Study Subjects	Controls	Relative Risks
Psychosocial problems	52.6 (23.2)†	1.3 (1.3)†	231 (63)†
Accidents, poisonings and violence	20.4	28.0	4.1
Urogenital problems	8.5	8.9	5.4
Digestive problems	7.2	8.9	4.5
Cardiovascular problems	4.2	26.1	0.9
Endocrine problems‡	2.3	1.9	6.7
Musculoskeletal problems	1.6	10.8	0.8
Neoplasms	0.9	—	—
Skin problems	0.8	0.7	7.0
Respiratory problems	0.6	1.9	1.7
Sensory/nervous problems	0.6	4.5	0.9
Symptoms and ill-defined conditions	0.3	3.2	0.6
Infectious diseases	—	3.2	—
Congenital anomalies	—	0.6	—
Total	100.0	100.0	
Weighted frequency scores§	879 (543)	157 (157)	5.6 (3.5)
Index of Dissimilarity: 52.7			

*See footnotes to Table 6 for explanations.

common than among controls. Urogenital problems and digestive problems represent 8 per cent and 7 per cent, respectively, of all causes of hospitalization for alcoholics. Alcoholics were at about five times higher risk of such hospitalizations than controls. Controls, on the other hand, were slightly more likely to receive inpatient services for musculoskeletal problems and for ill-defined conditions, while there is virtually no difference between the groups in reported cardiovascular and sensory or nervous problems. On the whole, alcoholics were nearly six times more apt to be hospitalized for all problems than controls. The index of dissimilarity for this distribution is 53 per cent.

The observed differences between study and control groups in the prevalence of psychosocial problems in both outpatient and inpatient service utilization are not attributable solely or even primarily to inflation of the figures by the overrepresentation of the problem code, alcoholism, among alcoholics. This is shown by the adjusted percentages and relative risks in parentheses in Tables 6 and 7. Strong group differences remain after exclusion of the "alcoholism" problem code. The prevalence of psychosocial problems among alcoholics is a real phenomenon, rather than an artifact of the coding.

Inappropriate Care-Seeking by Alcoholics

The finding that many outpatient noncounseling encounters by alcoholic study subjects involved psychosocial rather than strictly medical problems needs explaining. Since alcoholism figured prominently among alcoholics' psychosocial problems, some attention was clearly being paid to it. Primary-care providers at the HMO were directly addressing the alcoholism of about 40 per cent of study subjects, regardless of how it was made known to them. Knowledge may have come through the medical history, through contact with the spouse or other family members, through direct contact with an intoxicated subject, or through entries made in the medical records by counseling providers.

Based on their patterns of utilization, study subjects seem to have been burdening the medical-care delivery system by inappropriate care-seeking. Alcoholics tended to present in large numbers to non-counseling providers for problems that would have been more appositely handled by the counseling staff. Certainly, these alcoholics were not strangers to that department, having been selected on that basis.

What can happen when alcoholics are not treated for the primary problem of alcoholism is that providers may treat their symptoms and complaints in the same way that other anxious, depressed, nervous and fatigued patients may be treated, particularly by drug therapy. For patients' psychosocial problems, providers may prescribe certain central nervous system (CNS) depressants, including diazepam, oxycodone, chlordiazepoxide, and other drugs listed in Table 8. As they are alcohol-dependent, alcoholics may also become dependent on these prescription drugs. The *Physician's Desk Reference* (PDR) contains manufacturers' warnings that addiction-prone individuals, such as alcoholics, should be carefully followed in drug therapy with CNS depressants, because of the increased danger of drug overdose, due to the additive effects of these drugs with alcohol.[26]

In this sample, study subjects were much more likely than controls to be prescribed certain sedative, tranquilizing and other CNS depressant drugs during ambulatory non-counseling encounters. Table 8 shows the relative frequency with which alcoholics were prescribed a number of drugs that are potentiated by alcohol and carry manufacturers' warnings about the danger of overdosage and chemical dependency. In no case did controls receive more prescriptions for these drugs. Based on Table 8, alcoholics were about six times more likely to receive prescriptions for diazepam and chlordiazepoxide; they were about 12 times more likely to receive flurazepam and amitriptylene; and they were many times more likely to be given oxycodone and imipramine. Propoxyphene, which was three times more likely to be prescribed to alcoholics, is among the more dangerous of these selected drugs, if the PDR manufacturers' warnings are any indication.†

Diazepam was the most frequently prescribed drug, in this selected group of drugs, for both alcoholics and controls. Thirty-five alcoholics and 17 controls were prescribed the drug, according to medical record reports, during non-counseling encounters. This is not to say that the original prescriptions for certain drugs may not have come from the counseling department for some of their clients or that such drug prescriptions are not monitored and patients followed. The relatively high prevalence of drug overdoses and combined overdoses of alcohol and other drugs among hospitalized alcoholics, discussed below, should be viewed in light of these drug prescription patterns.

Related to the tendency of alcoholics to seek medical care inappropriately for emotional and psychosocial problems is the issue of the relationship of utilization behavior to length of time in counseling treatment. The question is whether study subjects who stayed in counseling treatment were the high utilizers of outpatient non-counseling services for psychosocial problems. This would be expected if high medical and high psychiatric utilization go together. On the other hand, the sample alcoholics who dropped out of counseling treatment might have been largely responsible for the prevalence of psychosocial problems reported or diagnosed by medical personnel. The subjects in this case would have been substituting primary medical care for counseling care. This and the related issue of whether HMO medical

---

† The *Physician's Desk Reference* (PDR) contains the warning, "Do not prescribe propoxyphene for patients who are suicidal or addiction-prone." Propoxyphene products ". . . in excessive doses, either alone or in combination with other C.N.S. depressants, including alcohol, are a major cause of drug-related deaths."[26]

TABLE 8. Numbers of Study and Control Subjects Receiving Prescriptions for Selected Sedatives, Anti-Depressants, Tranquilizers, and Other CNS Depressant Drugs During Ambulatory Non-Counseling Encounters and the Relative Risks Associated with Reports of Such Prescriptions

	Numbers of Subjects Receiving Prescriptions				
	Alcoholics		Controls		
Selected Drugs	Number of Cases*	WFS†	Number of Cases*	WFS†	Relative Risk
Diazepam	35	3255	17	544	6
Oxycodone	16	400	4	16	25
Flurazepam HCl	11	198	3	15	13
Imipramine HCl	11	187	0	0	—
Amitriptyline HCl	7	112	3	9	12
Chlordiazepoxide HCl	8	104	3	21	5
Propoxyphene	7	56	4	20	3

* Number of cases is the number of subjects for whom each drug was prescribed one or more times during outpatient encounters.

† WFS is the weighted frequency score, that is the number of subjects receiving each drug multiplied by the number of times each drug was reported as prescribed for all subjects combined.

services are used more appropriately and at reduced rates by alcoholics as a result of outpatient alcoholism treatment await further data analysis.

The Role of Accidents, Poisonings and Violence in Service Utilization

The predominance of psychosocial problems among alcoholics' inpatient problems masks other relationships. When psychosocial problem codes are excluded from alcoholics' inpatient diagnoses, the category of accidents, poisonings, and violence emerges as accounting for nearly half of all remaining inpatient problems. Urogenital and digestive problems also increase in importance in association with alcoholics' hospitalizations. These represent about one third of all inpatient problems, when the impact of psychosocial problems is eliminated.

That problems classified as accidents, poisonings and violence represent a larger share of both outpatient and inpatient problems for controls than for alcoholics requires some discussion. This is shown in Table 6. Accidents, poisonings and violence comprise 14 per cent of controls' outpatient problems, compared with 7 per cent of alcoholics' outpatient problems. Similarly, in Table 7, 28 per cent of all inpatient problems for controls were accidents, poisonings and violence, compared with 20 per cent of inpatient problems for alcoholics. This result is misleading, however, since comparison of the weighted frequency scores for the two groups shows that alcoholic study subjects were 20 per cent more likely to use outpatient services for accidents and related problems and were 300 per cent more likely to be hospitalized for such problems than were controls.

Table 9 shows the percentage distribution of particular categories of accidents, poisonings and violence, associated with inpatient utilization for alcoholics and controls. The two groups differ considerably in the problems that led most often to hospitalization. Fully 40 per cent of accidents, poisonings and violence associated with alcoholics' inpatient utilization were due to the adverse effects of medicinal agents. Nearly half were due to all types of adverse effects. By contrast, controls experienced *no* such problems. Lacerations and open wounds, especially of the head, neck and trunk, were frequent accidents associated with alcoholics' hospitalizations (16 per

TABLE 9. The Percentage Distribution of Problems Classified as Accidents, Poisonings and Violence Associated with Inpatient Utilization for Alcoholic Study Subjects and Controls

Types of Accidents, Poisonings and Violence	Percentage Distribution of Problems		
	Alcoholic Study Subjects	Controls	Relative Risk
Adverse effects of substances (medicinal and non-medicinal)	48.0	—	—
Lacerations and open wounds	15.7	56.8	1.1
Internal injuries*	10.6	2.3	19
Fractures	7.3	4.5	6.5
Superficial injuries	6.7	22.8	1.2
Contusions and crushing injuries	6.1	4.5	5.5
Sprains and strains of joints	5.0	2.3	9
Dislocations	0.6	4.5	0.5
Burns and effects of foreign bodies	—	2.3	—
Total	100.0	100.0	
Weighted frequency scores	179	44	4.1
Index of Dissimilarity: 63.4			

* This category includes intracranial injuries, internal injuries of the chest, abdomen, and pelvis, and injuries to the nerves and spinal cord.

cent), as were intracranial injuries, especially concussions (9 per cent), and fractures, especially of the skull, spine and trunk (7 per cent). Controls' most frequent accidents associated with inpatient utilization were lacerations and open wounds of the head, neck and trunk. These accounted for 57 per cent of all controls' accidents. Twenty-three per cent of their accidents were superficial injuries, compared with 7 per cent of alcoholics' accidents.

Clearly, the distribution of accidents for study subjects is powerfully skewed by the predominance of adverse substance effects. Within the adverse substance effects category, which accounts for such a high proportion of inpatient accidents and related problems among alcoholics, drug overdoses and combined alcohol and other drug overdoses are major contributors, constituting about 80 per cent.

When the relative risks for the two groups are examined, alcoholics were 6.5 times more likely to be hospitalized for fractures; 9 times more likely to be hospitalized for sprains and strains of joints; 19 times more likely to be hospitalized for internal, especially intracranial, injuries; and 5.5 times more likely to be hospitalized for contusions and crushing injuries, compared with controls. Both groups were about equally likely to be hospitalized for lacerations and superficial injuries. Controls, on the other hand, were slightly more likely to use hospital services for dislocations and for burns, although the numbers are very small.

The percentage distributions of particular accidents and related problems reported or diagnosed in conjunction with alcoholics' and controls' outpatient service utilization are given in Table 10. Of all accidents that occasioned outpatient encounters for study subjects, 59 per cent were sprains and strains of joints, 12 per cent were adverse effects of medicinal agents and other substances, 9 per cent were fractures, and 7 per cent were contusions and crushing injuries. Of controls' accidents associated with outpatient encounters, the vast majority (82 per cent) were sprains and strains of joints; 6 per cent were fractures and 4 per cent were adverse effects. There is more convergence between the distributions of accidents for which outpatient HMO services were utilized by alcoholics and controls than was the case for the inpatient comparison.

TABLE 10. The Percentage Distribution of Problems Classified as Accidents, Poisonings and Violence Associated with Ambulatory Encounters for Alcoholic Study Subjects and Controls

Types of Accidents, Poisonings and Violence	Percentage Distribution of Problems		
	Alcoholic Study Subjects	Controls	Relative Risk
Sprains and strains of joints	58.5	82.4	3.8
Adverse effects of substances (medicinal and non-medicinal)	12.2	3.9	3.8
Fractures	9.4	5.7	2.0
Contusions and crushing injuries	7.5	0.8	11.8
Lacerations and open wounds	6.0	3.8	1.9
Superficial injury	4.5	1.6	3.3
Dislocations	1.6	1.3	1.6
Internal injuries*	0.3	0.4	0.7
Burns and effects of foreign bodies	0	0.1	—
Total	100.0	100.0	
Weighted frequency scores	2040	1698	1.20
Index of Dissimilarity: 24			

* This category includes intracranial injuries, internal injuries of the chest, abdomen and pelvis, and injuries to the nerves and spinal cord.

This is shown by comparing the index of dissimilarity from Table 9—63 per cent—with that of Table 10—24 per cent.

Nevertheless, here again, in nearly every problem area, alcoholic study subjects utilized outpatient services for more accidents, poisonings and violence than did controls. The former sought outpatient care for twice as many fractures (2.5 times more skull, spine and back fractures); about twice as many lacerations and open wounds; 3 times more superficial injuries; 12 times more contusions; and 4 times more adverse effects of drugs and other substances. Controls, on the other hand, utilized outpatient services for 15 per cent more sprains and strains of joints and for slightly more internal injuries, specifically intracranial injuries, and more burns and effects of foreign bodies. Again, the numbers in these last categories are very small.

The fact that alcoholics were four times more likely than controls to be hospitalized for accidents, poisonings and violence and only about 20 per cent more likely to utilize outpatient services for such problems is puzzling. One possible explanation for this is that alcoholics may be going to hospital emergency rooms with problems that controls might take to the HMO. This is not likely, since the decision to seek hospital emergency room care is not usually made by individual HMO members, but is normally made by RIGHA providers, often by telephone. This process of permission-seeking is designed to reduce or eliminate inappropriate and uneven utilization of emergency room services. It insures that given similar problems, both study subjects and controls would be hospitalized at the same rate.

Another possible explanation is that the type of accidents and related problems that alcoholics have tend to be more serious and life-threatening than those experienced by controls. They are more likely therefore to require inpatient services. This notion is supported by the fact that "adverse effects," particularly drug and alcohol overdoses, were so frequently associated with inpatient encounters for alcoholic study subjects and *never* led to inpatient utilization by controls. The synergistic and potentiating effects of alcohol on other CNS depressant drugs are probably major causal factors here.[5,27] Such adverse effects were also associated with four times more outpatient utilization for alcoholics than for controls. This suggests an important way in which distinctly alcohol-related and alcohol-induced problems represent a burden on this health-care delivery system.

It is likely that alcoholics are reluctant to seek care for minor injuries for a number of reasons: because they do not want their alcoholism to be discovered in the process of explaining what happened; because being anesthetized and sedated by alcohol, they may hardly notice such injuries; and because being able to mobilize one's faculties and resources to seek medical care may be delayed by intoxication. By the time "morning comes," a laceration may be too old to suture, and the motivation to seek care may recede until a complication of the injury, such as infection, arises (Dr. LeClair Bissell, personal communication, August 14, 1980).

Two types of accidents do not appear to be directly or clearly alcohol-related. One is lacerations and open wounds, especially of the head, neck and trunk, for which alcoholics and controls were equally likely to be hospitalized. The other is sprains and strains of joints, for which controls were actually more likely to utilize outpatient services than were study subjects. Since many of the controls who suffered lacerations may have been involved in automobile and other accidents with alcoholics or under the influence of alcohol themselves, the role of alcohol here is difficult to determine. However, as noted, alcoholics may ignore many relatively minor accidental injuries for which outpatient care might be sought by others. Thus, the surprisingly high outpatient utilization by controls for sprains and strains may be attributable, in large part, to controls' tendencies to attend to such relatively minor injuries, before they are exacerbated or self-limited. Finally, it should be noted that alcoholics were hospitalized nine times more often than controls for sprains and strains, lending support to the idea that when alcoholics are injured and seek treatment for their injuries, their condition is more serious, and the services required are correspondingly more extensive and costly. In addition, if controls are generally healthier than alcoholics and if controls are less likely to utilize HMO services for psychosocial problems and for chronic and serious illnesses than are alcoholics, then it is understandable that accidents and related problems represent a higher proportion of problems associated with controls' service utilization.

There is debate over whether accidents, poisonings and violence are directly associated with alcohol, or whether, as some writers suggest, certain personality characteristics or some antecedent conditions give rise to both alcoholism and accident-proneness.[5] More research is needed in this area, but it seems likely that a combination of the anesthetic, sedative effects of alcohol, the possibility of potentiation of these effects, when alcohol is used with other CNS depressants, and the behavior associated with intoxication is central to explaining many types of accidental injuries, both unintentional and deliberate.[27]

Chronic Illness Among Alcoholics

A final brief insight may be gained into the issue of whether alcoholics utilize more HMO services, partly because they are more likely to be chronically or seriously ill than controls. Tables 11 and 12 provide lists of the most frequent problems, based on single, non-collapsed ICD codes, associated with outpatient and inpatient encounters, for comparison between study and control groups. They give some indication that chronic problems are likely to predominate among those most frequently reported for alcoholic study subjects. More acute problems seem to be listed among those most frequently recorded for controls.

Setting aside consideration of emotional problems, alcoholics tended to report more frequent chronic conditions, such as, diabetes, hypertension and obesity, than controls, judging by the magnitude of rela-

TABLE II. The Most Frequent Problems Associated with Outpatient Non-Counseling Encounters for Alcoholics and Controls and the Corresponding Relative Risks

Alcoholics				Controls			
Problems in Order of Frequency	No. Cases*	WFS†	RR‡	Problems in Order of Frequency	No. Cases*	WFS†	RR‡
1. Alcoholism	36	3492	175	1. URI§	31	1395	—
2. Hypertension	24	3096	3	2. Sprains and strains of sacroiliac	19	1140	5.7
3. Nervousness/fatigue	37	2331	6	3. Hypertension	12	1044	—
4. Obesity	27	1728	2	4. Obesity	19	855	—
5. Depression	31	1581	12.3	5. Allergic rhinitis	16	544	1.1
6. URI§	31	1519	1.1	6. Anxiety	16	400	—
7. Anxiety	25	1425	3.6	7. Nervousness/fatigue	15	390	—
8. Diabetes	9	756	10.1	8. Bursitis or tendonitis	12	384	1.3
9. Sprains and strains of back and neck	15	615	4.5	9. Costochondritis, etc.	9	306	—
10. Costochondritis, osteochondritis, epicondylitis	13	598	2.0	10. Low back pain	10	290	0.74

* Number of individuals reporting one or more occurrences of the problem.
† Weighted frequency score.
‡ RR = Relative Risk.
§ URI = Upper respiratory tract infection.

TABLE 12. The Most Frequent Problems Associated with Inpatient Encounters or Hospitalizations of Alcoholics and Controls

Alcoholics					Controls		
	Problems in Order of Frequency	No. Cases	WFS*		Problems in Order of Frequency	No. Cases	WFS*
1.	Alcoholism	12	336	1.	Face or scalp lacerations	5	25
2.	Drug overdose	6	54	2.	Acute MI, heart attack	3	18
3.	Depression	5	40	3.	Angina pectoris	3	12
4.	Pregnancy	5	35	4.	Ruptured disc	2	12
5.	Face or scalp lacerations	5	25	5.	Pregnancy	3	9
6.	Suicidal ideation	3	24	6.	Gallbladder problems (cholecystitis, cholelithiasis)	4	8
7.	Diabetes	3	18	7.	Superficial injury of face, neck, or head	2	6
8.	Obesity	4	16	8.	Strep throat	2	4
	Fatty metamorphosis of the liver, hepatomegaly	4	16		Epilepsy, seizure disorder	2	4
	Female sterilization	4	16		Arterial embolism/thrombosis	2	4
	Concussion	4	16		Inguinal hernia	2	4
					Myalgia	2	4

* WFS is the Weighted Frequency Score, defined as the total number of occurrences of the problem code weighted by the number of cases or individuals whose records contained at least one mention of that problem.

ive risks for outpatient problems, shown in Table 11. Among the problems diagnosed most frequently for controls were such acute problems as upper respiratory tract infections (URI), sprains and strains of the sacroiliac region, and allergic rhinitis.

This conclusion is tenuous, partly because it is difficult to draw unambiguous boundaries between acute and chronic, and mild and serious conditions, and partly because obesity and hypertension were frequent problems for controls receiving outpatient care, as upper respiratory tract infection (URI) was a frequent problem for study subjects. These study subjects share many of the diagnoses listed as most frequent for alcoholics by Boyajy and by other authors.

Regarding hospitalizations, alcohol-related and/or chronic conditions seem to predominate among the most frequent problems for alcoholics (Table 12). Besides alcoholism itself, these include drug overdose and depression, and chronic conditions, such as, diabetes, fatty metamorphosis of the liver/hepatomegaly, and obesity. The controls' list again tends to include more acute problems among those most frequently associated with inpatient utilization. These conditions are strep throat, face and scalp lacerations, other injuries to the face, neck and head, ruptured disc, and acute myocardial infarction (MI). Anomalous conditions, such as angina and gallbladder problems, are also among controls' most frequent problems. Face and scalp lacerations and pregnancy, which was coded as a problem, are frequently associated with hospitalizations for both groups. If these single problem codes could be combined into meaningful categories to approximate degrees of severity or chronicity, a better test of this hypothesis would be possible.

Conclusions

Summary

The purpose of this study was to compare the utilization of HMO services by two groups of HMO members—a group of 106 identified alcoholics and an equal number of matched controls. Utilization and other data were collected from HMO medical records. Utilization was measured by numbers and proportions of encounters compared across four service utilization categories—laboratory and x-ray services, external referrals, hospitalizations and ambulatory non-counseling encounters. Differences between study and control groups in the distribution and frequency of problems or morbidities associated with each major utilization category were examined.

The results showed that, as measured by encounters, alcoholics utilized more of all HMO services studied than did controls. Study subjects were especially prone to utilize inpatient hospital services, even when hospitalizations for alcoholism detoxification and rehabilitation were excluded.

Some evidence was found for the notion that alcoholics utilize more HMO inpatient and outpatient services, because they are more likely to be physically and emotionally ill than controls.

In the first place, more problems were identified for alcoholic study subjects, compared with controls, in association with all types of encounters, in particular, with hospitalizations.

Secondly, significantly more of the outpatient non-counseling and the inpatient problems reported by the alcoholic study subjects were classified as psychosocial, compared with controls' problems. The importance of accidents, poisonings and violence increased as causes of hospital service utilization by alcoholics, when psychosocial problem codes were excluded. Alcoholics were found to utilize more outpatient and inpatient services than controls for such problems, although accidents represented a larger proportion of all problems for controls. Alcoholics were much more likely to be hospitalized for drug overdoses than controls. In general, study subjects seemed to seek treatment for injuries, when their condition was more serious and when the services required to treat them were more extensive and costly, than was the case for controls.

Finally, some tentative suggestions were made, based on a comparison of the most frequent outpatient and inpatient problems for alcoholics and controls. These support the notion that more serious or chronic problems tended to be associated with alcoholics' inpatient and ambulatory encounters. By contrast, more acute problems were noted for controls.

Further Research Needed

The second stage of this research project will address the issue of whether and to what extent the failure to engage alcoholic HMO members in counseling treatment is costly to the health care delivery system, viewed in terms of their propensity to use a disproportionate share of HMO services. How acceptance of alcoholism treatment affects service utilization patterns and the associated illness experience of alcoholics will also be studied. One question of interest, for instance, is whether study subjects' utilization of medical services for hypertension, diabetes, obesity and other alcohol-related problems, as well as for psychosocial problems, such as anxiety and depression, will decrease with treatment. Some of the policy implications of the results of offering third-party coverage for alcoholism treatment, particularly in HMOs will be discussed.

An investigation of the possible "dose-response-type" relationship of alcoholism and morbidity is also of interest. Using measures of the level and extent of alcohol involvement or impairment, including the duration of excessive alcohol consumption, Boyajy's contention may be tested. Do more severely involved alcoholics experience higher morbidity and use more services than other alcoholics?

A third area of vital concern is sex differences in alcoholics' morbidity and care-seeking. The general issue is whether the illness of alcoholism might act as a leveler of expected sex differences in service utilization. Male alcoholics may be found to utilize health care services at the typically high rates of females generally, excluding obstetric-gynecologic and other sex-specific problems and procedures.[28] If male and female alcoholics do utilize services at similar rates, then the question persists whether and how patterns of HMO use and associated morbidities may differ by sex. As mentioned earlier, how alcoholic women differ from their controls in utilization behavior and illness experience is an important aspect of this research issue.

The question of causality remains. How alcoholism affects the origin and/or the development of certain illnesses and injuries needs clarification. Some combination of factors seems to be operating here, including the physiologic effects of alcohol ingestion; possible alcohol-related conditions and habits, such as inadequate nutrition, exposure to the cold, lack of physical exercise, general neglect of health, and increased susceptibility or lowered resistance to infections[6]; combined alcohol and other drug use, including coffee or tea consumption, cigarette smoking, the use of tranquilizers, sedatives, and stimulants[11]; background or personality factors, including genetic predisposition; and social factors, such as anomie, social isolation, etc.[29] All of these variables may interact with alcohol and confound its effects on illness. Psychiatric problems, such as nervousness, anxiety, and depression, of which alcoholics have been shown to be at significantly higher risk than controls, may, according to Pell and D'Alonzo, ". . . lead to psychosomatic disorders and increased susceptibility to a wide range of diseases,"[6] especially those linked to stress. The issue is very complex, and relationships are difficult to disentangle. However, determination of the extent to which the excessive consumption of alcohol in and of itself is important in the natural history of certain illnesses, independent of other risk factors, is critical to our understanding of alcoholism and the other illnesses that may signal it.

Acknowledgments

The author gratefully acknowledges the expert technical assistance of Dr. Robert L. Stout, Assistant Professor (Research), Section of Psychiatry and Human Behavior, Brown University, Providence, Rhode Island and the generous advice and useful comments of Dr. LeClair Bissell, President and Chief Executive, Edgehill Newport, Newport, Rhode Island. Thanks are also due to Michael C. Gerhardt, the Occupational Alcoholism Project Director at Rhode Island Group Health Association, for his early conceptualization of the research problem and his editing suggestions, to the therapeutic staff of the counseling department and to the medical records staff for their cooperation, and to the staff of the Employee Assistance Program, especially to Wendy Woifson, for her most able and skillful research assistance, and to Sandra Peah, for typing the manuscript.

References

1. Berry R, et al. The economic costs of alcohol abuse and alcoholism—1977. Prepared for the National Institute on Alcohol Abuse and Alcoholism. Rockville: PB 278-606 AS.

2. JWK International Corporation. Benefit-cost analysis of alcoholism treatment centers. Volumes I and II. Prepared for the National Institute on Alcohol Abuse and Alcoholism (#ADM-281-75-0031) Annandale: 1976.

3. Deakin CML, Santamaria JN, Wilkinson P. The cost of alcoholism to the community. Med J Aust 1973;i:1305.

4. Wersinger R, Roberts J, Roghmann K, Wells S. The inpatient hospital and ambulatory utilization experience of an alcoholic population compared to a matched control group within an HMO. Presented at the twenty-eighth annual meeting of the Group of Health Institute, New York, June 18-21, 1978.

5. United States Department of Health, Education and Welfare. Alcohol-related illnesses. First Special Report to the U.S. Congress on Alcohol and Health. From the Secretary of Health, Education and Welfare. Rockville: U.S. Department of Health, Education and Welfare, 1971:45 (DHEW Publication #ADM 74-68).

6. Pell S, D'Alonzo CA. The prevalence of chronic disease among problem drinkers. Arch Envir Health 1968;16:679.

7. Barchha R, Stewart MA, Guze SB. The prevalence of alcoholism among general hospital ward patients. Am J Psychiatry 1968;125:681.

8. Wilkinson P, Kornaczewski A, Rankin JG, Santamaria JN. Physical disease in alcoholism—initial survey of 1,000 patients. Med J Aust 1971;i:1217.

9. Moore RA. The prevalence of alcoholism in a community general hospital. Am J Psychiatry 1971;128:638.

10. Brody JA, Mills GS. On considering alcohol as a risk factor in specific diseases. Am J Epidemiol 1978;107:462.

11. Williams AT, Burns FH, Morey S. Prevalence of alcoholism in a Sydney teaching hospital: some aspects. Med J Aust 1978;ii:608.

12. Penick EC, Norton JC, Sauers CA, Powell BJ, Read MR. Medical problems of the alcoholic in nonmedical treatment. Health Soc Work 1978;3:123.

13. DeLint J, Schmidt W. Alcoholism and mortality. In: Kissin B, Begleiter H, eds. Social aspects of alcoholism. New York: Plenum Press, 1976:275.

14. Pell S, D'Alonzo CA. A five-year mortality study of alcoholics. J Occup Med 1973;15:120.

15. Massé H. L'alcoolisme facteur de mortalité: son incidence sur les principales causes de décès. La Nouvelle Presse Méd 1972;1:1857.

16. Brock P, Coordinator. Summary. Report of the liaison task force on alcohol-related problems. Submitted to the President's Commission on Mental Health, February 15, 1978 (PCMH/P-78/26), 2078.

17. Putnam SL. Policy adoption—helps and hindrances: a study of factors associated with one measure of outcome success in an HMO-based occupational alcoholism project. Presented at the seventh annual meeting of the Association of Labor-Management Administrators and Consultants on Alcoholism, San Diego, October 3-6, 1978:217.

18. Boyajy TG. Alcoholism within prepaid group practice. Washington, D.C.: Group Health Association of America, 1978.

19. Boyajy TG. Demographic, patient functioning, and health care utilization profiles of alcoholics' pre-post treatment within three prepaid group practices. Presented at the twenty-eighth annual meeting of the Group Health Institute, New York, June 18-21, 1978.

20. Wan TT, Siofer SJ. Determinants of physician utilization: a causal analysis. J Health Soc Behav 1974;15:100.

21. McKinlay JB. Some approaches and problems in the study of the use of services—an overview. J Health Soc Behav 1972;13:115.

22. Anderson OW, Andersen RM. Patterns of use of health services. In: Levine S, Reeder LG, eds. Handbook of medical sociology. 2nd ed. Englewood Cliffs: Prentice-Hall, 1972;386.

23. Blue Shield of Rhode Island. Schedule of relative values for surgery; schedule of x-ray procedures; schedule of diagnostic laboratory services. Blue Shield of Rhode Island Manual, Providence, no date, 25.

24. Committee on Nomenclature and Statistics. Detailed list of major disease categories in ICD-8. DSM-II—Diagnostic and statistical manual of mental disorders. 2nd ed. Washington, DC: The American Psychiatric Association, 1968;83.

25. Blalock HM Jr. Social statistics. 2nd ed. New York, McGraw-Hill Book Company, 1972:223.

26. Baker CE Jr, Publisher. Physician's desk reference (PDR). 34th ed. Oradell, N.J.: Medical Economics Company, 1980.

27. Waller JA. Alcohol and unintentional injury. In: Kissin B, Begleiter H, eds. Social aspects of alcoholism. New York: Plenum Press, 1976;307.

28. Ashley MJ, Olin JS, le Riche WH, Kornaczewski A, Schmidt W, Rankin JG. Morbidity in alcoholics: evidence for accelerated development of physical disease in women. Arch Intern Med 1977;137:883.

29. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979;109:186.

-1 C 5170-1 -

Medical Care

February 1982
Vol. 20 No. 2
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Physician Performance in a Prepaid Health Plan: Results of the Peer Review Program of the Health Insurance Plan of Greater New York

Episode-Based Efficiency Comparisons for Physicians and Nurse Practitioners

Measurement of Task Delegations Among Nurses by Nominal Group Process Analysis

Costs and Outcomes of Medical Intensive Care

Predicting the Outcome of Primary Care

Hospital Discharge Queues in Massachusetts

Privacy and Computerized Health Information in Community Health Centers and District Health Agencies

EPSDT Impact on Health Status: Estimates Based on Secondary Analysis of Administratively Generated Data

PUBLISHED BY
J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $45.00, special price U.S. students, $23.00, all other countries except Japan, one year, $55.00. Single copies $6.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A.
Copyright © 1982 by J. B. Lippincott Company.
Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

February 1982
Volume 20, Number 2

CONTENTS

ORIGINAL ARTICLES

Physician Performance in a Prepaid Health Plan: Results of the Peer Review Program of the Health Insurance Plan of Greater New York. *Jeanne Magagna Deuschle, Barbara Alvarez, Donald N. Logsdon, William M. Stahl and Harry Smith, Jr.* 127

Episode-Based Efficiency Comparisons for Physicians and Nurse Practitioners. *David S. Salkever, Elizabeth A. Skinner, Donald M. Steinwachs and Harvey Katz* 143

Measurement of Task Delegations Among Nurses by Nominal Group Process Analysis. *Vandan M. Trivedi* 154

Costs and Outcomes of Medical Intensive Care. *Mark R. Chassin* 165

Predicting the Outcome of Primary Care. *Diana Dryer Wright and Robert L. Kane* 180

Hospital Discharge Queues in Massachusetts. *Leonard W. Gruenberg and Thomas R. Willemain* 188

Privacy and Computerized Health Information in Community Health Centers and District Health Agencies. *Barry Trute and Robb Tonn* 202

EPSDT Impact on Health Status: Estimates Based on Secondary Analysis of Administratively Generated Data. *Patrick H. Irwin and Rosemary Conroy-Hughes* 216

COMMUNICATION

Primary Care and Medical Education. *Michael Ross and Barry Willer* 235

BOOK REVIEWS

Hospital Costs and Health Insurance, by Martin Feldstein. *Alain C. Enthoven* 242

Continuing Education in the Health Professions, by Robert Boissoneau. *Tricia C. Corbett* 246

Perspectives on Canadian Health and Social Services Policy: History and Emerging Trends, edited by Carl A. Meilicke and Janet L. Storch. *Eugene Vayda and Elaine Vayda* 248

Poverty and Health: Economic Causes and Consequences of Health Problems, by Harold S. Luft. *Richard M. Scheffler* 249

PUBLICATIONS RECEIVED 251

MEDICAL CARE
February 1982, Vol. XX, No. 2

Original Articles

Physician Performance in a Prepaid Health Plan:

Results of the Peer Review Program of the Health Insurance Plan of Greater New York

JEANNE MAGAGNA DEUSCHLE, M.A.,* BARBARA ALVAREZ, M.A.,† DONALD N. LOGSDON, M.D., M.P.H.,‡ WILLIAM M. STAHL, M.D.,§ AND HARRY SMITH, JR., PH.D.‖

A peer review program to evaluate the quality of medical care was established by the Health Insurance Plan of Greater New York in 1973. Physician performance is assessed through application of explicit process criteria to medical care as recorded in the patient record. A total of 6,788 records were reviewed in terms of the clinical management of acute otitis media, hypertension and breast lesions. Follow-up procedures to bring about positive changes in delivery of health care were integrated into the program at the time of the initial audit. Readit of 715 records was used to measure change in physician compliance with medical care standards. Results of this audit/reaudit process demonstrated statistically significant improvement in quality of clinical practice as measured by the assessment parameters.

NUMEROUS STUDIES have been conducted on the quality of health care in prepaid group practices and health maintenance organizations (HMOs). Cunningham and Williamson recently reviewed and evaluated literature on the quality of health care in an HMO as compared with that in other settings with similar populations.[1] The primary focus of these studies was the assessment of the quality of care. As expected, there were wide variations in the definitions of quality, the methodologies used and the criteria that were applied for the measurements. This article reports the changes demonstrated by a Peer Review Program audit and reaudit of ambulatory health care in a prepaid group practice HMO.

Background

The Health Insurance Plan of Greater New York (HIP) is one of the oldest and largest prepaid group practice health insurance plans in the country. In 1979, there were approximately 775,000 members receiving comprehensive health care on a prepaid basis through contractual arrangements that involved 26 multispecialty medical groups with 48 ambulatory centers in the metropolitan New York City area. HIP was approved as a New-York-State–certified HMO in 1979.

Numerous studies have been conducted at HIP on the quality of health care, includ-

---

* Director, Quality Assurance Division (QAD), Health Insurance Plan of Greater New York (HIP), New York, New York.

† Biostatistician, HIP, QAD.

‡ Director, Lifecycle Preventive Health Services Project, Sponsored by INSURE, New York, New York.

§ Professor of Surgery, New York Medical College, New York, New York.

‖ Chairman, Department of Biomathematical Sciences, Mount Sinai School of Medicine, New York, New York.

Reprint Requests: Jeanne M. Deuschle, Director Quality Assurance Division, Health Insurance Plan of Greater New York, 220 W. 58th St., New York, NY 10019.

0025-7079/82/0200/0127/$01.30 © J. B. Lippincott Co.

ing the early (1951) Makeover report,[2] the comparison studies by Shapiro on prematurity and perinatal mortality (1958 and 1960),[3,4] and patterns of medical care for an elderly, indigent population (1967).[5]

The current Peer Review Program at HIP was established in 1973 for the purpose of measuring the quality of medical care in each of the 28 (subsequently merged to 26) affiliated medical groups and to identify and resolve problems in the delivery of ambulatory health care by group physicians. The results reported in this paper are those of a large operational program with many practical methodologic constraints that were imposed because of the complex nature of the prepaid health system in which it was developed. The tracer methodology, as described by Kessner,[6] was utilized at HIP; the implementation of this methodology has already been reported.[7]

The first phase of the HIP evaluation process is based on the comparison of initial audit and reaudit results. This approach resulted in large-scale reviews (7,503 patient records) over a period of four years, involving multiple delivery sites—with proper sampling design, quality control for data collection, standardized analysis, presentation and follow-up of results.

Methods

Initial Audit Procedure

The HIP Peer Review Program assesses physician performance through application of explicit process criteria to medical care for selected tracer[6] conditions, as recorded in patient records.

For the first system assessment, medical care standards were established for hypertension in adults, acute otitis media in children and breast lesions. Criteria for each of these tracer conditions were formulated by a peer physician committee representing HIP and affiliated medical groups. Other areas for which medical care standards were subsequently developed are: urinary tract infections in adults, cervical neoplasia, anemia in children, acute pharyngitis in children, and follow-up of selected abnormal laboratory test values. The medical care standards are reviewed periodically for necessary modifications based on changes in medical management.

The basis for patient selection is the HIP Physician Visit Report Form. Historical information was obtained on the actual number of services provided by each of the 28 medical groups in the specialty areas involved, the total enrollment for each and estimate of the prevalence of the tracer condition. The size of the sample of patient's records for each medical group was determined by a simple random sampling formula, using an estimate of the standard deviation of quality scores within a group obtained from a pretest trial and establishing a 95 per cent confidence interval on the true group quality mean score of 10 quality points. If the prevalence figure for a group resulted in a small potential group population of patients for assessment, finite correction factors were used.

Data required for application of criteria are collected from patient records by professionally trained and supervised lay abstractors, using specially constructed data collection instruments.#

Following analysis (which utilized item importance weights[8] and is discussed in detail in another paper[7]), the Peer Review Program staff presents tables of results to the Peer Review Committee for discussion and formulation of recommendations. Medical groups are not identified in the presentations, except by code. This policy of confidentiality was established in an effort to ensure the Program's educational objectives. A Peer Review Committee report, which includes recommendations for

---

Careful analysis of the time required to complete the audits was made to ensure optimal utilization of manpower. The median time for review of a hypertension case was 32 minutes, for an acute otitis media case, 12 minutes, and for a breast lesion case, 20 minutes.

corrective action, item compliance and total median compliance, is then sent to the appropriate medical group director.

Members of the peer review staff, including a physician consultant, meet with the medical group's physician medical audit committee representatives for an in-depth discussion of the report and recommendations of the Peer Review Committee. In many instances, specific suggestions are made to assist the group in addressing identified problems.

The medical group director is required to respond in writing to the recommendations of the Peer Review Committee within two months. This response must include the medical group's plan to improve physician compliance with the medical care standards and, in particular, to address problem case listings that accompany the report. The plan of most medical groups included use of follow-up meetings with group physicians and/or conduct of mini-audits. In addition, changes in administrative procedures were sometimes used to facilitate resolution of problems.

Reaudit Procedure

The Peer Review Committee decided that a period of two months was adequate to permit medical groups to implement a physician education effort and to correct those problem areas identified in the reports and recommendations of the Peer Review Committee. In order to determine to what extent improvement had been made throughout the HIP system, a reaudit procedure was established, using the same medical care standards as for the initial audit.

Selection of Patient Records. Cases selected for reaudit were those diagnosed a minimum of two months after discussion of initial review results with medical group physicians.

Since it was desirable to assess improvement in physician compliance with medical care standards as soon as possible, it was decided to select a sample size large enough to estimate the true percentage of compliance within ± 5 percentage points, with a risk equal to 5 per cent at the system level (the 28 medical groups as an aggregate), rather than to estimate compliance for each of the medical groups. This resulted in a reaudit of a small number of cases (n = 13) from each medical group, with a total of n = 364 for the system, providing an adequate assessment of system improvement in physician compliance with medical care standards.

Data. For reaudit of acute otitis media, data were collected and analyzed in a manner identical to that of the original audit. However, for reaudit of hypertension, the data collection method was modified and was based on a combination of newly diagnosed and continuing care cases. Those newly diagnosed were collected and analyzed as in the original audit. The continuing care cases were selected randomly from those included in the initial audit. The collection and analysis of these cases were limited to assessment of compliance with standards of care for 1) longterm follow-up visits and 2) recording of blood pressure at each follow-up visit.

Of the 13 hypertension cases sampled in each medical group, nine were continuing care, and four were newly diagnosed. The smaller proportion of those newly diagnosed was decided upon because of the difficulty in identifying new cases of hypertension within a short period of time.

A systemwide reaudit for breast lesions was not indicated because of uniformly high physician compliance with medical care standards.

Reporting of Reaudit Results and Medical Group Follow-up. The reporting system described for presentation of initial audit data was also used for reaudit reporting, with minor modifications. For example, initial audit compliance was added to each table to facilitate comparison with reaudit and subsequent preparation of the Peer Review Committee report and recommendations. As was done for the initial audit, peer review staff met with physician medical audit committee representatives of each medical group to discuss the reaudit report. For medical groups with continued "unsatisfactory" physician compliance with medical care standards, an additional group full-sample reaudit was conducted (as described for initial audit). The Peer Review Committee considered unsatisfactory performance as a median total compliance below 60 per cent on both audit and reaudit, as well as a statistically significant decrease in compliance on reaudit, regardless of the group's audit compliance.

Outcome analyses were considered in planning the scope of the peer review approach, but because of the complexity and limited strategies available to assess outcomes in operational settings, it was decided to limit initial efforts to review of the medical process. The collected process data do, however, include some intermediate outcome results. For example, for hypertension, blood pressure levels are recorded for all follow-up visits. Since we had not controlled for medical and

TABLE 1. Peer Review Program, Acute Otitis Media—26 Medical Groups, Comparison of Audit and Reaudit, Number and Per Cent of Cases in Which Physician Complied With Each Item Collected, Median Weighted Compliance by Category; Number of Cases in Initial Audit: 2106; Number of Cases Reaudited: 352

	Weight	Audit		Reaudit	
		Number	Per Cent	Number	Per Cent
Diagnostic Process	25%				
History					
Description of symptoms	5%	1422	67.5	289	82.3*
History of specific allergies	3%	793	37.7	205	58.4*
Physical					
Temperature	1%	581	27.6	157	44.6*
Tympanic membrane	5%	1226	58.2	292	83.2*
Cervical lymph nodes	3%	360	17.1	123	34.9*
Pharynx/tonsils	3%	1041	49.4	223	63.4*
Laboratory					
Culture and sensitivity	5%	17	24.6	7	41.2
Number of Applicable Cases[†]		(69)		(17)	
Therapeutic Process	75%				
Treatment					
Acceptable antimicrobial drug prescribed on diagnosis date	33.3%	1942	92.3	324	92.0
Adequate dosage recorded	16.7%	1166	61.5	230	72.6*
Follow-up					
Request return visit in 3–21 days	12.5%	1314	62.6	259	73.6*
Further evaluation	12.5%	62	73.8	8	53.3
Number of Applicable Cases[‡]		(84)		(15)	
Median Compliance					
Total		70.2		83.3*	
Diagnostic Process		50.4		70.0*	
Therapeutic Process		77.7		100.0*	
Grid Area 5—Median Total Compliance		50.0		66.9*	

* Statistically significant difference, $p < 0.05$.

[†] Applicable only if record of draining ear.

[‡] Applicable if within the 12-month period prior to this episode child had three or more episodes or six or more visits for otitis media, or a recording of suspected hearing loss stated.

patient-related factors that also influence outcome, we have not presented these outcome results.

Results

Acute Otitis Media

Initial Audit Results. The 2,106 cases of acute otitis media (criteria for application of medical care standards are presented in Appendix A) reviewed in the initial audit of the HIP system were diagnosed between 1972 and 1977.

Table 1 presents the number and percentage of cases in which the physicians complied with each item and the median weighted compliance by category for the HIP system. In general, items in the diagnostic process, with 25 points of weight, were less frequently recorded than those in the therapeutic process, with 75 points of weight.

The median total compliance for the 2,106 cases was 70.2 per cent compliance for diagnostic process, 50.4 per cent, and for therapeutic process, 77.7 per cent. There was a great deal of variation in item compliance among the medical groups. For example, a recording concerning the status of the tympanic membrane, with 58.2 per cent overall compliance, ranged from 4–100 per cent among the individual medical groups. There was 92.3 per cent compliance for the item on acceptable antimicrobial drug prescribed on diagnosis date, with little variation among the medical groups. In most instances, noncompliance was attributed to no drug order being recorded in the patient record rather than to the prescription of an unacceptable drug. For 5.6 per cent, no drug was recorded on the date of diagnosis, and for only 1.4 per cent, the drug was unacceptable. Compliance with the drug item was not required if a physician impression of serous/viral otitis media was recorded.

There was 61.5 per cent compliance with the item "adequate dosage recorded." Noncompliance for this item was for the most part attributable to an absent/incomplete recording of dosage (34.1 per cent). For the remaining 4.3 per cent the dosage prescribed was judged to be below optimal levels.

A return visit in 3–21 days, with 62.6 per cent compliance, includes physician request for a revisit, regardless of whether the patient complied.

Figure 1 shows the median total compliance and the percentage of cases with total compliance of 50.0 per cent or higher for acute otitis media for the HIP system, as well as the distribution of the 28 medical groups for these parameters. The grid was originally developed as a management tool, so that improvement efforts could be focused on those medical groups falling into the lowest grid categories. This is discussed in detail in another article.[7]

The HIP system median compliance for acute otitis media was 70.2 per cent (Fig. 1). This tracer condition showed the greatest spread of the three for individual medical group compliance, which ranged for this condition from 33.5 per cent to 96.5 per cent.

Reaudit Results. The 352 cases of acute otitis media that were reaudited were diagnosed between 1976 and 1979.

Table 1 presents a comparison of audit and reaudit scores for individual items assessed, as well as the median compliance by category. Using the Mann-Whitney test,[9] statistically significant improvement ($p < 0.05$) was demonstrated for all median categories—total, increased 70.2–83.3 per cent; diagnostic process, 50.4–70.0 per cent; therapeutic process, 77.7–100.0 per cent.

Those items deemed most important to the medical management of the patient showed statistically significant improvement (binomial approximation to the normal, $p < 0.05$) or remained at the same high level of compliance as on initial audit, for example, prescribing an acceptable antimicrobial drug, 92.3 per cent on audit and 92.0 per cent on reaudit. Description of

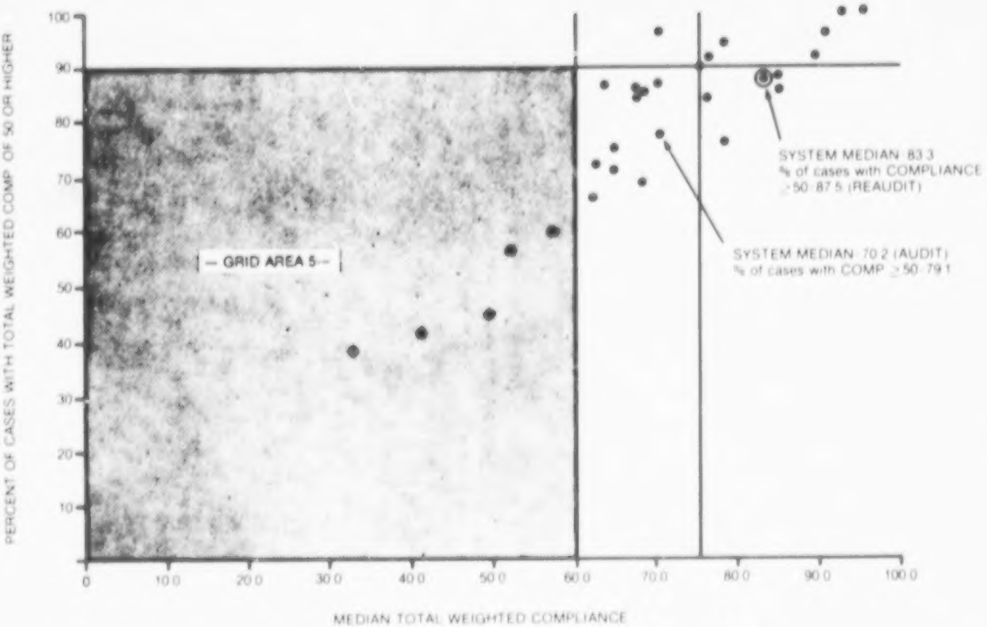


FIG. 1. Acute otitis media. Comparison of audit and reaudit (new cases); criteria for determining relationship of performance—28 HIP medical groups.

tympanic membrane on physical examination showed significant improvement—58.2 per cent on audit, and 83.2 per cent on reaudit. Compliance with the items on adequate drug dosage and a follow-up visit showed significant improvement. It is of interest that the noncompliance for adequate dosage was of the same kind as that found on initial audit, i.e., primarily absent or incomplete recording.

One item, further evaluation for those patients who had recurrent otitis media or hearing loss, decreased, but the result is not significant.

It is of particular importance that the median compliance for the aggregate of cases in the medical groups with audit compliance in the lowest grid category (see shaded area of Fig. 1) showed statistically significant improvement (Table 1)—50.0 per cent on audit to 66.9 per cent on reaudit (Mann-Whitney, $p < 0.05$).

Hypertension

Initial Audit Results. The 2,868 cases of hypertension (criteria for application of medical care standards are presented in Appendix A) reviewed in the initial audit of the HIP system were diagnosed between 1972 and 1977. Table 2 presents the number of cases in which the physicians complied with each item and the median compliance by category for the HIP system. Ninety per cent of the cases were classified as mild hypertension (diastolic 90–114 mm, systolic 140 mm or greater); 8 per cent as moderate hypertension (diastolic 115–130 mm); and 2 per cent severe hypertension (diastolic greater than 130 mm). In those instances in which more than one blood pressure was recorded at a visit, the reading with the lowest diastolic blood pressure was used. The median total compliance was 63.6 per cent; diagnostic process compliance was 55.0 per cent and

TABLE 2. Hypertension—26 Medical Groups, Comparison of Audit and Reaudit—New Cases; Number and Per Cent of Cases in Which Physician Complied With Each Item Collected
Median Weighted Compliance by Category; Number of Cases in Initial Audit: 2868*,
Number of Cases Reaudited: 111†

	Weight	Audit		Reaudit	
		No.	%	No.	%
Diagnostic Process	40%				
Disease Detection	4%				
A blood pressure recorded within 2 years prior to diagnosis		1316	61.5	81	84.4‡
Number of Applicable Cases§		(2139)		(96)	
History	14.0				
Personal/social history	2.0	1603	56.2	83	75.5‡
Family history	4.0	1441	50.5	71	64.5‡
History of previous diagnosis/treatment of hypertension	4.0	1510	52.9	94	85.5‡
History of related symptoms	4.0	1483	52.0	74	67.3‡
Physical	7.0				
A blood pressure recorded within 2 months after diagnosis	1.0	1872	65.8	83	77.6‡
Weight	1.0	1625	56.9	79	72.5‡
Funduscopic examination	2.0	581	20.4	25	24.0
Heart examination	1.0	2128	74.4	97	88.2‡
Abdominal description	1.0	1414	49.5	67	62.0‡
Extremities	1.0	1292	45.3	64	58.7‡
Laboratory and Other Tests	15.0				
Urinalysis	3.0	1888	66.0	83	77.6‡
Hematocrit or hemoglobin	1.0	1868	65.3	76	71.0
Blood urea nitrogen or serum creatinine	2.0	1597	55.9	76	71.0‡
Electrolytes—Serum potassium (prior to administration of antihypertensive drug therapy)	1.0	518	20.8	26	32.9‡
Blood sugar	1.0	1781	62.3	84	78.5‡
Serum cholesterol	1.0	1544	54.0	76	71.0‡
Electrocardiogram	3.0	1492	52.3	62	57.9
Chest x-ray	3.0	1742	61.0	81	74.3‡
Median Compliance					
Total		63.6		75.0‡	
Diagnostic Process		55.0		69.5‡	
Therapeutic Process		70.0		77.7‡	

	Weight	Audit		Reaudit	
		No.	%	No.	%
Therapeutic Process	60%				
Treatment					
Initial treatment regime		2147	74.9	98	88.3‡
Mild hypertension—5 week status	5.0				
Moderate hypertension—1 week status	6.0				
Severe hypertension—immediate status	30.0				
Six-month treatment/response status					
Mild hypertension only	5.0	378	60.4	17	63.0
Number of Applicable Cases		(626)		(27)	
Initial acceptable antihypertensive drug prescribed at acceptable dosage		1812	73.3	70	88.6‡
Mild hypertension	10.0				
Moderate hypertension	12.0				
Two-month response to drug therapy OR alteration of drug regime		1606	69.5	54	76.1
Mild hypertension	5.0				
Moderate hypertension	6.0				
Referral to consultant/hospitalization/diastolic blood pressure reduced to 130 mm or less within 1 month of date of diagnosis					
Severe hypertension only	30.0	47	75.8	—	—
Number of Applicable Cases		(62)		(0)	
Serum potassium ordered after antihypertensive therapy initiated		226	9.3	7	10.0
Mild hypertension	2.5				
Moderate hypertension	3.0				
Potassium replacement ordered		73	21.4	1	—¶
Mild hypertension	2.5				
Moderate hypertension	3.0				
Number of Applicable Cases #		(341)			
Follow up (Mild and Moderate Only)					
Per Cent of Cases in Full Compliance					
Initial follow up	10.0	1272	44.8	57	51.8
Long-term follow up	10.0	1373	51.9	66	75.0‡
Blood pressure recorded at follow-up visits	10.0	2224	78.4	100	90.9‡
Grid Area 5—Median Total Compliance		58.8		76.6‡	

---

* In audit, 90.1% of cases reviewed were classified as mild hypertension, 7.6% as moderate, and 2.3% as severe.

† In reaudit, 93.7% of cases reviewed were classified as mild hypertension and 6.3% as moderate.

‡ Statistically significant difference, $p < 0.05$.

§ Applicable to patients enrolled in the medical group for at least 2 years prior to detection of hypertension.

¶ Percentage not reported; number of applicable cases less than 10.

that for therapeutic process was 70.0 per cent.

As was found with acute otitis media, there was wide variation in item compliance among the medical groups. For example, a recording concerning family history was found in 50.5 per cent of all cases, but the individual medical group compliance ranged from 21–89 per cent.

Compliance for funduscopic examination was 20.4 per cent, with a range of 1–38 per cent among the medical groups. This item, which was the most heavily weighted physical examination item in the diagnostic process, was the most poorly recorded. Ordering of serum potassium (20.8 per cent prior to administration of antihypertensive drug therapy) was another item low in physician compliance, ranging from 5–45 per cent among the individual medical groups. In general, there was good compliance with the initial therapeutic regimen[¶] (74.9 per cent).

Physician compliance with the item concerning initial acceptable antihypertensive drug at acceptable dosage was also high (73.3 per cent), with a range of 56–90 per cent. Noncompliance with this item was as follows: antihypertensive drug or dosage not acceptable, 7.4 per cent; dosage incompletely recorded or illegible, 19.3 per cent. Compliance with the two-month response status to antihypertensive drug therapy was 69.5 per cent.[**]

There was full compliance with the initial follow-up standards[††] in 44.8 per cent of the cases (i.e., all expected visits were requested by the physician), and 51.9 per cent of the cases were in full compliance with long-term follow-up standards.

Figure 2 shows that while there was some degree of spread among the medical groups, the majority clustered in the lowest grid areas. The range of medical group compliance was from 57.3–74.0 per cent.

Reaudit Results. The 111 new cases of hypertension that were reaudited were diagnosed between 1977 and 1979.

Table 2 presents a comparison of audit and reaudit compliance for individual items, as well as median compliance by category. Statistically significant improvement (Mann-Whitney, $p < 0.05$) was demonstrated for all median compliance categories: total, 63.6 per cent to 75.0 per cent; diagnostic process, 55.0 per cent to 69.5 per cent; therapeutic process, 70.0 per cent to 77.7 per cent.

As was shown for acute otitis media, those items deemed most important in terms of medical management of the patient either showed statistically significant improvement, using the normal approximation to the binomial, $p < 0.05$, or remained at the same high level of compliance found on initial audit. For example, in newly diagnosed cases of hypertension, both an acceptable initial treatment regime (74.9 per cent on audit to 88.3 per cent on reaudit) and acceptable antihypertensive drug/dosage (73.3 per cent on audit to 88.6 per cent on reaudit) showed siginificant improvement. Incomplete or nonrecording of drug/dosage was the principal reason for physician noncompliance with this item. There are two important items that, although showing some improvement, could improve further: 1) six-month re-

---

[¶] *Initial treatment regime:* For cases classified as *mild* hypertension, any or all of the following were required with one month of index blood pressure reading: a follow-up visit, diet advice, sedation and antihypertensive drug treatment within one month of blood pressure reading. For cases classified as *moderate* hypertension, initiation of antihypertensive drug therapy was necessary within one week of index blood pressure reading. For cases classified as *severe* hypertension,immediate initiation of antihypertensive drug therapy or referral to consultant or hospitalization was required.

[**] *Two month response status* required 1) the diastolic blood pressure to be reduced by 10 mm; if only systolic hypertension, the systolic blood pressure to be reduced by 10 mm unless physician stated that such reduction was contraindicated; and 2) if no response by blood pressure reduction, then drug dosage to be modified.

[††] *Follow-up standards required: Initial follow up*–at least one visit every two months for four months for cases classified as mild hypertension and one visit every month for three months for cases classified as moderate; *Long-term follow up*–one visit every six months for both mild and moderate hypertension.

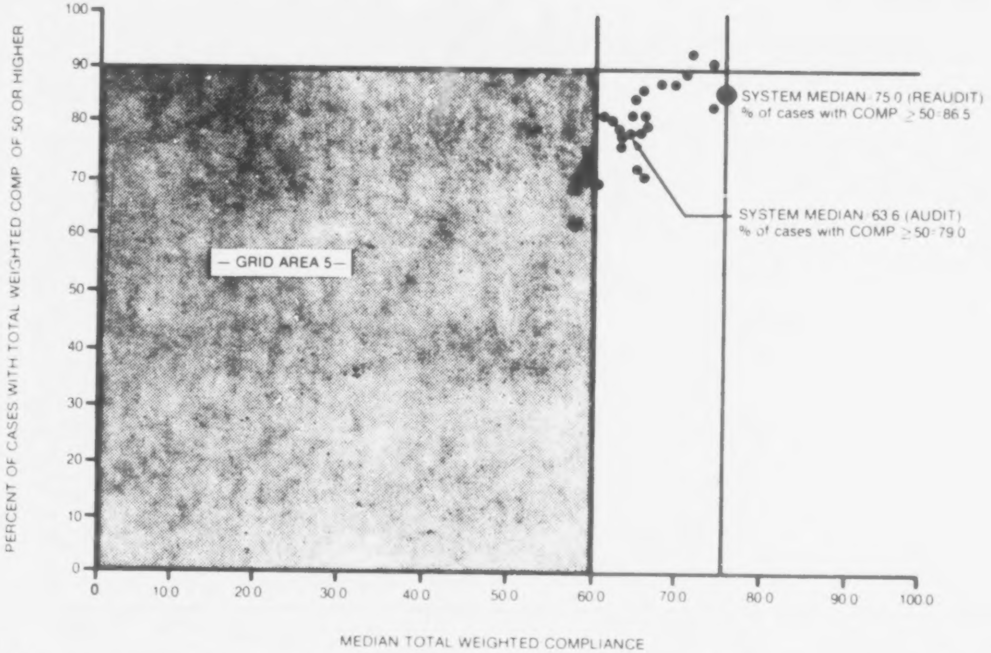


FIG. 2. Hypertension. Comparison of audit and reaudit (new cases); criteria for determining relationship of performance—28 HIP medical groups.

sponse to treatment for patients with mild hypertension who were not initially placed on an antihypertensive drug regime [six-month response status (for patients with mild hypertension who were not initially placed on an antihypertensive drug regime) required reduction of blood pressure to normal range within six months or initiation of antihypertensive drug therapy] and 2) full compliance with standards for initial follow-up visits.

Of the 19 items included in the diagnostic process, 16 showed statistically significant improvement. Physician compliance with medical standards for funduscopic examination and pre- and posttreatment electrolytes remained exceptionally low.

As was found for acute otitis media, the median total compliance for the aggregate of cases for those medical groups with audit compliance in the lowest grid category (see Fig. 2, shaded area) showed statistically significant improvement (Table 2)–58.8 per cent on audit to 76.6 per cent on reaudit (Mann-Whitney $p < 0.05$).

A sample of cases from those included in the initial audit was also reaudited for hypertension in order to assess the status of continuing care (Table 3). Two items were reaudited: 1) physician compliance with medical care standards for long-term follow-up where improvement was measured in terms of full compliance with all expected follow-up visits and 2) recording of blood pressure at all visits. Two hundred fifty-two cases included in the initial audit were reaudited. Statistically significant improvement (McNemar test,[9] $p < 0.05$) in physician compliance with medical standards was demonstrated for both long-term follow-up (48.1 per cent on audit to 63.6 per cent on reaudit) and recording of blood pressure at each visit (77.9 per cent on audit to 90.6 per cent on reaudit).

TABLE 3. Peer Review Program, Hypertension—26 Medical Groups, Comparison of Audit and Reaudit—Continuing Care (n = 252 Cases)

	Audit		Reaudit	
	No.	%	No.	%
Cases in Full Compliance				
Long-term follow-up	111	48.1	147	63.6*
Number of Applicable Cases†	(231)		(231)	
Blood pressure recorded at visits	166	77.9	193	90.6*
Number of Applicable Cases†	(213)		(213)	

* Statistically significant difference, $p < 0.05$.

† The method used derived 252 cases, of which 231 had sufficient time (i.e., 6 months) to assess compliance with at least one long-term follow-up visit. Of these 231 cases, only 213 had at least one visit, thereby allowing a blood pressure recording to be assessed.

Breast Lesions

Initial Audit Results. The 1,814 cases of breast lesions reviewed in the initial audit of the HIP system were diagnosed between 1968 and 1977.

The cases were classified as either benign or suspicious, based on the physician's statement describing the lesion and the diagnostic impression as recorded in the patient's record. Only cases with a diagnostic impression of benign breast disease, e.g., fibrocystic disease, were classified as benign. All others, e.g., mass, lump, were classified as suspicious. A surgeon's diagnosis was given priority over a diagnosis by a family physician. In the occasional instances for which it was not possible for the lay abstractor to classify a case, it was reviewed by the Program's physician consultant. In 1,108 cases (61.1 per cent), the lesions were classified as benign, and in 706 cases (38.9 per cent), they were classified as suspicious (Table 4).

Physician compliance for breast lesions was much higher than for acute otitis media and hypertension. The median total compliance was 87.5 per cent. Compliance for the diagnostic process was 73.3 per cent and for the therapeutic process, compliance was 100.0 per cent.

For the diagnostic process, physician compliance in recording the history items was low, with essentially little difference among the medical groups. The three physical examination items assigned a total weight of 17 points were in general well performed, in particular, description of the breast lesion (89.6 per cent), which required for compliance that the physician record the size, location and right/left breast. Physician compliance with the ordering of mammography was high (88.4 per cent). This item was applicable only to those women who were in follow-up for the breast lesion for a minimum of two years.

Physician compliance with the therapeutic process standards for breast lesions classified as benign required only follow-up visits (initially every three months for six months, then every six months), with a description of the lesion at each visit. There were 500 cases (45.1 per cent) in full compliance with this standard. This means that all expected visits, as well as a description of the lesion, were recorded.

Physician compliance with medical care standards for lesions classified as suspicious required a biopsy within two months. If a biopsy was performed between two and six months, partial compliance was assigned. There was very high physician compliance with this standard—597 cases (84.6 per cent).

Of the 703 cases classified as "suspicious" (three cases deleted from suspicious

TABLE 4. Peer Review Program, Breast Lesion Results—28 Medical Groups—Initial Audit, Number and Per Cent of Cases in Which Physician Complied With Each Item Collected; Median Weighted Compliance by Category
(Number of Cases Reviewed: 1814)

	Weight	Number	Per Cent
Diagnostic Process	30%		
Disease Detection			
Breast examination within 2 years prior to diagnosis of initial breast lesion	6%	742	59.7
Number of Applicable Cases*†		(1242)	
History			
Previous breast disease or breast surgery	1%	120	6.6
Hormone therapy	1%	185	10.2
Menstrual history	1%	418	23.0
Physical			
Examination of both breasts	4%	1291	71.2
Regional lymph nodes	3%	862	47.5
Description of breast lesion (size, location, and right/left breast)	10%	1626	89.6
Other			
Mammography	4%	1017	88.4
Number of Applicable Cases*‖		(1151)	

	Weight Assigned to Each Type	Number and Per Cent in Each Category	
		Number	Per Cent
Therapeutic Process			
Total Weight	70%		
Benign lesion‡ (No biopsy required)	0–70%		
Number and per cent of cases in *Full*§ (100%) compliance with follow-up standards		500	45.1
Suspicious lesion‡ (Biopsy required)	0–70%		
Number and per cent of cases in *Full*§ (100%) compliance with therapeutic process standards		597	84.6
Median Compliance	Weight		
Total	100%	87.5	
Diagnostic Process	30%	73.3	
Therapeutic Process	70%	100.0	

* The number of applicable cases is only listed for items on which the denominator of the per cent differs markedly from the total number reviewed.

† Applicable to those patients who were over 35 years of age and enrolled in the medical group for at least two years prior to detection of breast lesion.

‡ Number and per cent of lesions classified as: benign lesions: n = 1108; % = 61.1;
suspicious lesions: n = 706; % = 38.9.

§ Full compliance for benign lesions = follow-up visits with complete description of lesion at frequency specified in standards.
Full compliance for suspicious lesions = as for benign and biopsy within two months.

‖ Applicable if performed or to those cases that had two full years of review available.

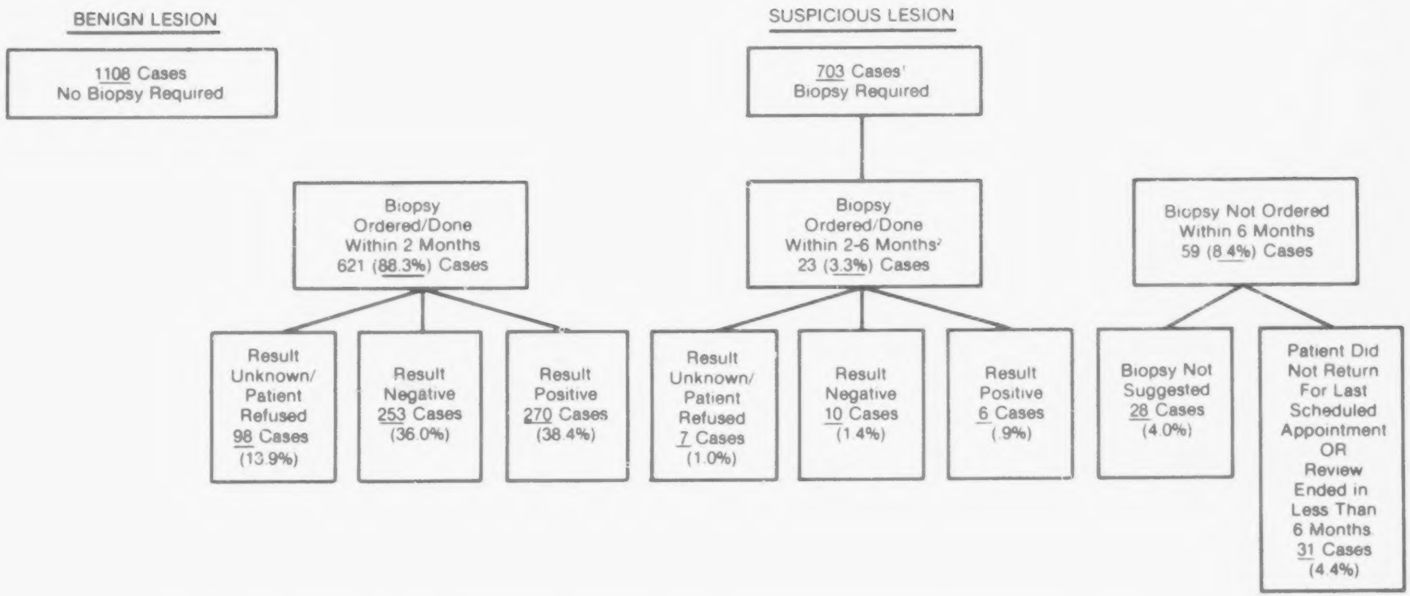

BENIGN LESION
1108 Cases
No Biopsy Required
Biopsy
Ordered/Done
Within 2 Months
621 (88.3%) Cases
Result
Unknown/
Patient
Refused
98 Cases
(13.9%)
Result
Negative
253 Cases
(36.0%)
Result
Positive
270 Cases
(38.4%)
SUSPICIOUS LESION
703 Cases[1]
Biopsy Required
Biopsy
Ordered/Done
Within 2-6 Months[2]
23 (3.3%) Cases
Result
Unknown/
Patient
Refused
7 Cases
(1.0%)
Result
Negative
10 Cases
(1.4%)
Result
Positive
6 Cases
(.9%)
Biopsy Not Ordered
Within 6 Months
59 (8.4%) Cases
Biopsy Not
Suggested
28 Cases
(4.0%)
Patient Did
Not Return
For Last
Scheduled
Appointment
OR
Review
Ended in
Less Than
6 Months
31 Cases
(4.4%)

FIG. 3. Peer Review Program. Flow chart of results of breast lesion cases reviewed—HIP system (n = 1,814).

---

[1] There were three additional cases classified initially as suspicious by a surgeon. Lesion not present at subsequent examination by surgeon.

[2] In a few instances, the time period may actually be up to two years if last visit in an interval is a broken appointment.

classification—see Fig. 3), biopsy was suggested/ordered within two months for 621 cases (88.3 per cent) and within two to six months for 23 (3.3 per cent) (Fig. 3).

Of the three tracer conditions reviewed, breast lesions showed the least spread in compliance, with all but one medical group in the highest grid category (Fig. 4).

Reliability of Audit and Reaudit Data

Procedures were implemented in the ongoing system to measure and assure the reliability and accuracy of the data collected. Following these procedures, the field supervisor, on a daily basis, selected a random sample (approximately 5 per cent) of the patient records that had already been abstracted and resubmitted them to be read a second time by another abstractor. In addition, a subsample of these duplicate records was selected periodically and read a third time by the field supervisor.

Data obtained from both audit and reaudit resulted in a reliability rate for abstracted items exceeding 90 per cent, consistent with the reproducibility rate for lay abstractors of medical records reported by Lyons and Payne[8,10] and Hulka.[11]

Discussion

As Brook has pointed out, "No new conceptual frameworks for assessing quality have been developed during the last two decades." Assessment of structure, process and outcome variables are all used but, as yet, ". . . there is no evidence to indicate which one produces the most valid judgments of the program or practice being assessed . . . opinion differs as to which type of judgment is the most valid."[12]

Largely because of its feasibility in terms of ". . . objectivity, economy of data collection, and adaptability to computer analysis, as compared to assessment of outcomes or the use of physician reviewers,"[13] HIP's Peer Review Program implemented the most commonly used method to assess physician performance—the use of explicit process criteria for selected conditions.

Criteria or standards developed by peer consensus are compared with information available from the patient record. Christoffel has discussed limitations to this approach in detail.[14] A frequently cited limitation is concerned with problems encountered in the use of the medical record, primarily lack of documentation of medical management; examples include the common practice of nonrecording of "negatives" or normal findings, failure to record drugs and dosages, and failure to record requests for return visits—all of which can bias results to noncompliance.

In spite of the aforementioned problems, the authors are in agreement with Payne's postulate, ". . . that use of the medical record for measuring physician performance will improve communication, improve the quality of medical care, and improve the content of the medical record."[15] A positive relationship between good recording and medical care performance has already been demonstrated by Lyons and Payne.[16]

Related to this issue is the frequent comment that this type of peer review activity—application of explicit process criteria to the medical record—primarily addresses *recording* of medical care, and that improvement on reaudit indicates improved recording. Although our results do indicate that a significant amount of noncompliance is attributable to nonrecording, there are a number of examples of *recorded management* that were not in compliance with the standards. Several examples showing the distribution of noncompliance follow.

Acute Otitis Media—
System Audit

Noncompliance with the item "acceptable antimicrobial drug prescribed on diagnosis date" was 7.7 per cent. Presumably, most of this noncompliance was related to managerial rather than to recording prob-

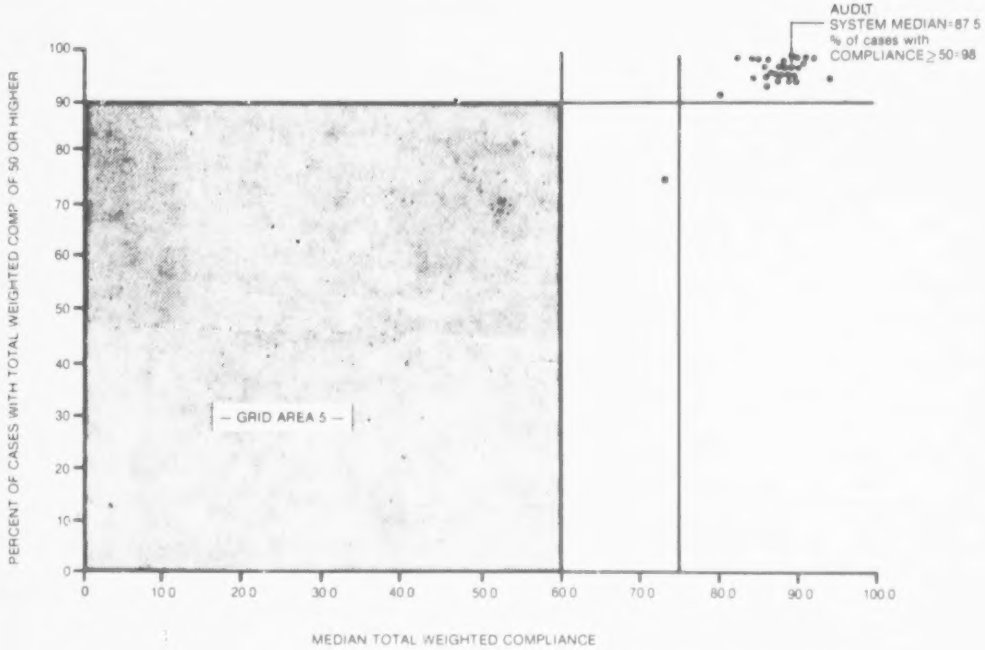


FIG. 4. Breast lesions. Criteria for determining relationship of performance—28 HIP medical groups.

lems, as follows: 5.6 per cent, no antimicrobial drug prescribed; 1.4 per cent, unacceptable antimicrobial drug prescribed (usually a cephalosporin; subsequent modification of the standards considered these drugs acceptable although not recommended); 0.6 per cent, penicillin prescribed in spite of physician's acknowledged statement of child's penicillin allergy. Noncompliance with this item remained unchanged on reaudit.

Noncompliance with the item "adequate dosage prescribed" was 38.4 per cent. While 34.1 per cent of this noncompliance was attributable to nonrecording of dosage, 4.3 per cent was because the dosage prescribed was judged to be below optional levels. Noncompliance with this item on reaudit decreased significantly, from 38.4 per cent to 27.5 per cent. The distribution of noncompliance showed impact on recording only.

Hypertension—System Audit

Noncompliance with the item "initial acceptable antihypertensive drug/dosage" was 26.7 per cent. While 19.3 per cent of this was attributable to nonrecording of drug dosage, 7.4 per cent was because of prescribing unacceptable drug or dosage. Essentially all of this was unacceptable dosage—only 0.5 per cent was unacceptable drug. It should be mentioned that dosage criteria did not accept a dosage prescribed for fewer than five days per week. Dosage criteria were made discretionary in a 1979 modification of these standards. Noncompliance with this item decreased significantly on reaudit, from 26.7 per cent to 11.4 per cent. The distribution of noncompliance showed impact both on recording and on prescribing of acceptable drug/dosage.

* * *

Figures 1, 2 and 4 show a significant difference in system compliance for the three conditions—acute otitis media, hypertension and breast lesions. This result implies that the system quality of medical care varies by condition, with breast lesions having the highest system compliance. It can be postulated that the high physician compliance with medical care standards for breast lesions may be related to characteristics of the condition (a malignant breast lesion produces greater morbidity than does hypertension or acute otitis media and, therefore, may be more carefully managed) or to characteristics of physicians in specialty practice (medicine and pediatrics versus surgery). Other factors that may also contribute to these differences in condition compliance are the number and kinds of items in each of the standards and the item weights assigned.

Brook has pointed out that "the most serious threat to the positive effect of quality assurance programs on care and health status lies with failure to use the results of quality assessment studies to improve care. . . . Effective quality assurance programs must take full advantage of feedback mechanisms, such as the reaudit procedures . . . to ensure quality and improve health status."[12] From its inception, the HIP Peer Review Program has been aware of the importance of building into the assessment of physician performance a specific approach to assist medical group physicians in resolving identified problems. The results of audit/reaudit presented in this paper have demonstrated that a significant increase in physician compliance with the medical care standards was obtained following use of a feedback mechanism.

We fully recognize that causal statements cannot be made concerning the impact of the Peer Review Program on this improvement. Organizational constraints did not permit the development of the sophisticated research design that would be required to address this issue. However, HIP is encouraged by these results, and it is our impression that the activity of the Program provides a significant impetus for improvement in physician compliance with medical care standards.

In line with the continued need to find ways to use results for the improvement of medical care, the Peer Review Program is exploring mechanisms to encourage and assist the groups in the development of group-based quality assurance activities that can work in a complementary way with HIP's systemwide quality assurance program. It is our opinion that this change in quality assurance emphasis, from a system responsibility to a shared local responsibility with the medical groups, will be a catalyst to improved quality of medical care in the HIP system.

NOTE: Medical care standards and data collection methods are available from Mrs. Deuschle upon request.

References

1. Cunningham FC, Williamson JW. How does the quality of health care in HMOs compare to that in other settings? An analytic literature review: 1958 to 1979. The Group Health Journal 1980;1.

2. Makover HB. The quality of medical care, methodology of survey of the medical groups associated with the Health Insurance Plan of New York. Am J Public Health 1951;41:824.

3. Shapiro S, Weiner L, Densen PM. Comparison of prematurity and perinatal mortality in a general population and in the population of a prepaid group practice, medical care plan. Am J Public Health 1958;48:170.

4. Shapiro S, Jacobziner H, Densen PM, Weiner L. Further observations on prematurity and perinatal mortality in a general population and in the population of a prepaid group practice medical care plan. Am J Public Health 1960;50:1304.

5. Shapiro S, Williams JJ, Yerby AS, Densen PM. Patterns of medical use by the indigent aged under two systems of medical care. Am J Public Health 1967;57:784.

6. Kessner DM, Kalk CE. A Strategy for Evaluating Health Services. Contrasts in Health Status. Washington, D.C.: National Academy of Sciences, Institute of Medicine, 1973;2.

7. Deuschle JM, et al. Implementation of a peer review system for ambulatory care: Experience of the Health Insurance Plan of Greater New York. Public Health Rep 1978;93:258.

8. Lyons TF, Payne BC. The use of item importance weights in assessing physician performance with predetermined criteria indices. Med Care 1975;13:432.

9. Conover WJ. Practical Nonparametric Statistics. New York: John Wiley & Sons, 1971.

10. Payne BC, Lyons TF, et al. The Quality of Medical Care: Evaluation and Improvement. Health Services Monograph Series. Chicago, Ill.: Hospital Research and Educational Trust, 1976.

11. Hulka BS, et al. Peer review in ambulatory care: Use of explicit criteria and implicit judgments. Med Care 1979;17(suppl).

12. Brook RH, Davies-Avery A. Mechanisms for Assuring Quality of U.S. Medical Care Services: Past, Present, and Future. Prepared under a grant from Department of Health, Education, and Welfare, R-1939-HEW. Santa Monica: The Rand Corporation, 1977.

13. Rosenthal G. Foreward to NCHSR Research Summary Series. Quality of Ambulatory Care: An Exploration of The Discrepancy Between Explicit Process Criteria and Performance (Hulka). U.S. Department of Health, Education, and Welfare, Public Health Service, Office of Health Research, Statistics, and Technology, National Center for Health Services Research, 1980. [DHEW Publication No. (PHS) 80-3244.]

14. Christoffel T, Loewenthal M. Evaluating the quality of ambulatory health care: A review of emerging methods. Med Care 1977;15:877.

15. Payne BC. The medical record as a basis for assessing physician competence. Ann Intern Med 1979;91:623.

16. Lyons TF, Payne BC. The relationship of physicians' medical recording performance to their medical care performance. Med Care 1974;12:463.

Appendix A

1. Acute Otitis Media

The criteria used for application of the medical care standards were physician diagnosis of acute otitis media or, in the absence of physician diagnosis of acute otitis media on a patient's record, recording of painful ear, inflammation/erythema of tympanic membrane, diminished light reflex, fluid in middle ear cavity or bleb on tympanic membrane.

2. Hypertension

The criteria used for application of the medical care standards were a blood pressure reading at the following age and blood pressure levels:

Age	Systolic and/or Diastolic up to or Exceeding
18–44	140/90mm
45–64	150/95mm
65 and over	160/95mm

MEDICAL CARE
February 1982, Vol. XX, No. 2

Episode-Based Efficiency Comparisons for Physicians and Nurse Practitioners

DAVID S. SALKEVER, PH.D., ELIZABETH A. SKINNER, M.S.W.,
DONALD M. STEINWACHS, PH.D., AND HARVEY KATZ, M.D.

Most previous studies comparing the efficiency of new health practitioners with that of physicians have used the visit as the basic unit of output. Several researchers have noted, though, that the episode is a conceptually superior output unit in several respects, although it is more complex to deal with methodologically. This study demonstrates the application of episode-based methods for comparing the efficiency of physicians with that of nurse practitioners. Data are drawn from the information system of the Columbia Medical Plan and from observations of provider time inputs. The analysis is confined to care episodes for otitis media and sore throat in the Department of Pediatrics. Results indicate that per episode costs with nurse practitioners as the initial provider are approximately 20 per cent below the costs of episodes in which physicians are the initial provider. Examination of a limited amount of data on patient-reported measures of effectiveness indicates that while nurse practitioners' care is less costly, it is not less effective. These findings are particularly interesting in light of recent doubts expressed about cost-savings from using new health practitioners, and particularly nurse practitioners, in group practice settings.

A VARIETY OF APPROACHES have been used recently to compare the costs or productivity of physicians with that of new health practitioners (NHPs) in providing primary care. The most important difference among these approaches concerns the way in which the unit of output is defined. A number of studies have defined output in terms of a collection of specific types of tasks or visits. For example, Record et al.[1] divided all outpatient visits to the Kaiser-Portland internal medicine department into 23 categories and defined annual department outpatient output as the annual frequencies for each of these 23 types of visits. Costs of producing this output with physicians alone and with the least-cost combination of physicians and physician's assistants were then compared. An earlier study of primary care practices by Smith et al.[2] used 41 categories of visits to define output and examined the increase in output (i.e., the proportionate increase in visits of all types) that could be obtained by adding a physician's assistant to a practice.

Other studies using the visit as the basic unit of output have examined only one or several particular types of visits. Examples are the recent comparisons of well-baby care provided by pediatricians and pediatric nurse practitioners.[3–8] Still other analyses based on visits have not differentiated among visits by type or content of

---

From the Health Services Research and Development Center, The Johns Hopkins Medical Institutions, Baltimore, Maryland.

This research was supported by Grants HS00427 and HS01964 from the National Center for Health Services Research.

Address for Reprints: Elizabeth A. Skinner, M.S.W., Health Services R and D Center, The Johns Hopkins Medical Institutions, 624 N. Broadway, Baltimore, MD 21205.

0025-7079/82/0200/0143/$01.05 © J. B. Lippincott Co.

services. This is most common in before–after comparisons of cost per visit for practices that have added NHPs, or similar comparisons between practices with and without NHPs.[7,8] However, the productivity or efficiency implications of these comparisons are clouded by the possibility that the mix of visits is altered when NHPs are introduced into a practice.[9,10] If this mix changes, before–after or interpractice differences in cost per visit may be partly due to differences in the mix of visit types rather than to productivity differentials.

An alternative unit of output less frequently employed than the visit is the person-year of care (i.e., care of one person for a one-year period). This unit is only appropriate for situations in which defined comparison groups of patients are assigned either to a physician or to an NHP, who serves as the primary source of first-contact care and who initiates the referral process to other providers. Average cost per person per year for all services (or a defined subset of services) in each comparison group is used as the efficiency criterion. Examples are the evaluation by Chambers et al. of family practice services in rural Newfoundland[11] and the studies of chronically ill patients reported by Spector et al.[12] and Lewis and Resnik.[13] The most appealing feature of this type of cost comparison is that it is population-based. However, as already noted, it can only be applied in certain rather unusual circumstances. Moreover, such cost comparisons will clearly be sensitive to random variations in the incidence of health problems among comparison groups, so that even in a randomized trial results may be unreliable unless the comparison groups are quite large.

Another unit of output that has several conceptual advantages for cost and productivity comparisons is the episode of care. It has been pointed out that episode-based comparisons are consistent with the idea that treatment of health problems, rather than the provision of individual services, is the primary objective of medical care.[14,15] From this perspective, the episode is preferable to the visit as the unit of comparison because it takes into account differences in referral and revisit rates between physicians and NHPs. The episode is also a more appropriate unit for measuring differences in effectiveness of care, since the outcome of the care process may be causally related not only to service received at a single visit, but to any services received over the course of the episode. However, episode-based comparisons are methodologically more complex than comparisons based on visits, since procedures must be developed for defining episodes and identifying the services contained in an episode. Perhaps because of this added complexity, and in spite of their conceptual advantages, episode-based cost or productivity comparisons of physicians and NHPs have not been carried out previously. (A recent study by Moscovice does report episode-based comparisons between nurses and family nurse practitioners in treating otitis media and urinary tract infection.[16]) Research by Tompkins et al.[17] compared the cost-effectiveness of internists and algorithm-assisted physician's assistants in treating adult respiratory illnesses using a 10–14-day period as an approximation for the episode of care.

In this article we shall describe the methods and findings of efficiency comparisons between pediatricians and pediatric nurse practitioners in a group-practice HMO in treating episodes of otitis media and sore throat. A major objective of this research was to develop and demonstrate practical procedures for episode-based comparisons that maximize the use of routinely collected data and minimize interruption of patient and provider activities. It is suggested that such procedures make it feasible to carry out episode-based comparisons for a variety of types of health problems and in a variety of organized ambulatory care settings. The research reported here is also of substantive interest since the extent of cost or productivity gains from NHP employment in group practices is currently in doubt, with some analysts reporting large gains and others suggesting a much smaller impact.[18] Furthermore, it has been argued that existing evidence of cost or productivity gains for nurse practitioners is weaker than that for other types of NHPs.[18]

Study Setting

The site for the study was the Department of Pediatrics of the Columbia Medical Plan (CMP), an HMO serving approximately 23,000 enrollees in the city of Columbia, Maryland. The Department of Pediatrics was staffed by four "full-time equivalent" (FTE) physicians, four FTE nurse practitioners and approximately four health assistants. The department has employed nurse practitioners and health assistants since late 1969. The annual volume of patient visits is now approximately 24,000, with physicians being the primary provider in about 60 per cent of these visits and nurse practitioners being the primary provider in about 40 per cent of the visits.

Nurse practitioners and physicians both function as independent professionals in providing ambulatory care. Health assistants may perform a few well-defined specific clinical tasks (such as immunizations), but their primary role is in scheduling, telephone answering and clerical functions. They also carry out preliminary tasks during physician visits, such as bringing the patient and the patient's record to the examination room, weighing and temperature taking. At the time of this study, nurse practitioners usually performed these tasks for their own patients.

CMP's information system is composed of two parts. The enrollment file contains data on enrollees' demographic characteristics, while the encounter system maintains information on the nature and content of each visit. Data are entered from encounter forms completed at each visit and include patient name and enrollment number, appointment status, purpose of the visit, providers seen, laboratory, x-ray and other procedures ordered, problems or diagnoses, and disposition. Prescriptions filled by the plan's in-house pharmacy are entered as well. It is estimated that all but 5 per cent of prescriptions written by CMP physicians are captured by the system.[19]

Research Methods

As noted earlier, a primary objective of the research was to develop a relatively inexpensive procedure for estimating costs of episodes of care. To achieve this objective, our study plan incorporated findings from previous methodological research on the definition of episodes and relied heavily on the routinely collected data in the CMP information system. However, one important element in the cost estimation of ambulatory care episodes, which was not captured by the information system, was provider time inputs. (Duration of visit is recorded on the plan's encounter forms, but only in 15-minute intervals.) After consideration of alternative data collection methods, it was decided that the most economical and nonintrusive procedure for determining provider time inputs was to employ observers stationed outside of offices and treatment rooms who would record the comings and goings of providers and patients. This procedure was followed during a two-week period in May of 1976. In each observation period, observers were assigned to particular providers and recorded data on all visits to these providers. The observation schedule was designed to provide data on both walk-in and scheduled visits to each of the physicians and nurse practitioners in the Department of Pediatrics. Roughly half of all visits to the department during the two-week period were captured by the observers. (A more detailed description of the methods used in the observational study has been reported elsewhere.[20])

While the research focused on comparisons of episode costs, information was also available, from prior research at the CMP, on the quality and effectiveness of care provided by physicians and nurse practitioners. This was gathered as part of an Experimental Medical Care Review Organization (EMCRO) project. Data collection methods and analyses developed from this project have been reported previously.[21,22] In our research, further data analyses dealing specifically with otitis media and sore throat episodes were carried out.*

Definition and Selection of Episodes

Episodes of care for otitis media and sore throat were identified using the encounter data information system. Criteria were applied to link individual visits into episodes using diagnostic information, time interval, disposition and type of visit (initial illness, follow-up illness, etc.). A maximum interval of three weeks was allowed between successive visits for sore throat and purulent otitis media, while five weeks was used between serous otitis media visits. The selection of these intervals was based on clinical consensus, with comparisons between the information system and medical record to test the methodology.[23] The results indicate that descriptions of episodes of care derived from the information system using explicit criteria and those obtained through review of the medical chart are comparable. Exceptions are those cases in which otitis media is chronic and the distinction between individual episodes in not clear either in the chart or the information system. It should also be noted that episodes may involve more than one condition (e.g., otitis media and upper respiratory infection), which may be a significant determinant of longer episodes and the more intensive use of resources.

The episode definitions were applied to encounter system data for all visits to the Department of Pediatrics during the period of January through June, 1976. As a result, 409 otitis media episodes involving 646 visits, and 390 sore throat episodes involving 409 visits were identified.† These episodes did include a total of 17 visits to the CMP's Urgent Care Center, a walk-in clinic for after-hours services in which almost all care was rendered by physicians. However, no episodes had more than one Urgent Care visit, and all episodes where the initial visit was to Urgent Care were excluded from the study. Although it is possible that some visits to ENT specialists involved treatment of the two problems under study, it should be noted that linking these visits with specific episodes is problematic because ENT referrals usually are occasioned by repeated episodes or chronic problems. Using our method for defining episodes, only 20 ENT visits were identified in the 799 study episodes. Visits to other departments in the plan were presumably not related to otitis media or sore throat problems, and were not included in the episodes.

Time Inputs and Associated Costs

The costs of an episode of care were broken down into two components: 1) costs associated with provider time inputs and 2) costs of ancillary services and drugs. Costs associated with provider time inputs for

---

* While the EMCRO data on pediatric patients at the CMP were gathered two years prior to the other data used in our study, available evidence suggests that the characteristics of the patient population and practice patterns were essentially unchanged over this period.

† Note that episodes apparently beginning in early January or extending to late June could not be clearly delimited with the January–June data, based on our decision rules, they might in fact have begun at an earlier date or ended at a later date. Therefore, these episodes were excluded from the study.

TABLE 1. Mean Patient Contact Time Per Visit

	MD		NP	
	Time (min.)	N	Time (min.)	N
Otitis media				
Initial visits[*]	9.3	20	13.4	14
Follow-up visits	7.5	18	8.4	5[†]
Sore throat				
Initial visits[‡]	7.7	16	13.1	18
Follow-up visits	11.0	3	9.5	2

[*] MD-NP difference is significant at the 0.10 level (two-tailed).

[†] Excludes one six-minute visit with MD consultation.

[‡] MD-NP difference is significant at the 0.05 level (two-tailed).

the two types of episodes studied here were calculated on a per-visit basis for six categories of visits: initial and follow-up visits to physicians and to nurse practitioners with and without physician consultation. The first step in this calculation was to compute the average provider time spent in direct patient contact for each type of visit. These figures were based on the May 1976 observational study. From a total of 379 visits for which time data were obtained, 58 were for treatment of otitis media and 39 were for treatment of sore throat. The mean times for initial and follow-up visits by provider type are shown in Table 1. For both problem categories, physician initial visit times were considerably shorter, while physician and nurse practitioner times for follow-up visits in the data were similar. However, the numbers of follow-up visits in the data were very small, and only a single nurse practitioner visit with physician consultation (a follow-up otitis media visit) was recorded. Therefore, the cost calculations for nurse practitioner visits with physician consultations used the nurse practitioner times shown in Table 1 and assumed a physician contact time of 4.6 minutes. (The 4.6 minute figure was the average physician contact time for all nurse practitioner visits with physician consultation, regardless of problem type, recorded in the observational study.) Furthermore, because of the limited numbers of otitis media and sore throat visits observed, the sensitivity of study findings of changes in mean time values was also examined and is discussed below.

Noncontact provider time inputs for each category of visit were assumed to be proportional to direct contact time. The proportionality factors used were the ratio of noncontact time to contact time during walk-in hours observed in the two-week study. Data for walk-in hours rather than scheduled-visit hours were used because 90 of the 97 observed otitis media and sore throat visits occurred during walk-in hours. These ratios were 0.84 for physicians and 0.73 for nurse practitioners.

For visits including physician time inputs, the services of health assistants must also be considered. Since detailed data on health assistants were not obtained in the observational study, we assigned 0.5 minutes of health assistants' time for each minute of physician time. The 0.5 figure was based on the fact that one health assistant was assigned to assist two physicians in their patient care activities, while other health assistants carried out scheduling, telephone and clerical duties.

The final component of costs associated with provider time inputs is space costs. Each pediatrician utilized 280 square feet of space in one office and two examining rooms, while each nurse practitioner was assigned an office with 120 square feet of space. Assuming that space was utilized for 2,000 hours annually and employing an annual cost figure (supplied by the plan's financial office) of $9.50 per square foot, space costs per minute of physician and nurse practitioner time were figured at 2.22 cents and 0.95 cents, respectively.

In summary, the formula for provider-time-related costs at each visit is:

$$(T_{MD} + 0.84T_{MD})\ (W_{MD} + 0.5W_{HA} + 2.22) + (T_{NP} + 0.73T_{NP})\ (W_{NP} + 0.95),$$

where T_{MD} and T_{NP} are the minutes of the physician and nurse practitioner direct contact time, and where W_{MD}, W_{HA}, and W_{NP} are the salary and fringe benefit costs per minute for physicians, health assistants and nurse practitioners. The plan's financial data indicated values of 40 cents for W_{MD}, 8 cents for W_{HA} and 15 cents for W_{NP}. Inserting these values and the T_{MD} and T_{NP} figures reported into this formula yielded the cost figures shown in Table 2. (Note that $T_{MD} = 0$ for nurse practitioner visits without consultation and $T_{NP} = 0$ for physician visits.) The follow-up physician visit cost figure in Table 2 was also applied to the 17 follow-up visits in the Urgent Care Center. Costs for the 20 ENT visits were not included in our calculations. Since 13 of these visits occurred in episodes in which the physician was the initial provider, this omission tends to bias our cost comparisons in favor of physicians.

TABLE 2. Provider-Time–Related Costs per Visit

	MD	NP	NP with MD Consultation
Otitis media			
Initial	$7.91	$3.70	$7.61
Follow-up	$6.38	$2.32	$6.23
Sore throat			
Inital	$6.55	$3.61	$7.52
Follow-up	$9.35	$2.62	$6.53

Ancillary Service and Drug Costs

The cost of drugs was figured as the acquisition costs to the plan, plus a $2 service charge. This is the formula used by the plan in its own billing system. Similarly, costs for x-ray services were equated with the rates billed for these services to fee-for-service customers. Laboratory costs were equal to charges paid by the plan to outside laboratory facilities.

Episode Cost Calculations

The costs for each episode were computed by adding the relevant cost figure in Table 2 for each visit in the episode to the cost of ancillary services and drugs. For example, the cost of an otitis media episode consisting of an initial nurse practitioner visit, two follow-up physician visits, and $5.00 in drug and ancillary costs would be $3.70 + 2 ($6.38) + $5.00 = $21.46. Mean cost figures for episodes grouped by problem type and provider type at the initial visit were then calculated.‡

Analysis of EMCRO Data

As part of the Columbia EMCRO project,[21] questionnaire data on patient-reported problem status were gathered for all visits to the Department of Pediatrics during a two-week period in mid-1974. Problem status was described in terms of amount and intensity of pain and discomfort in the previous 24 hours, degree of activity limitation in the previous three days and level of anxiety about the problem. One week telephone follow-ups were also made for a 50 per cent random sample of patients to determine change in problem status subsequent to the visit. Tabulations of these data for persons with initial visits for otitis media and sore throat are presented below. The purpose of these tabulations was to determine whether patient-reported problem severity and changes in problem status differed between physician and nurse practitioner patients.

---

‡ Standard deviations of the mean cost figures were also calculated and used in the statistical tests reported. However, these standard deviations do not precisely measure (and probably understate) the variability in episode costs for several reasons. They only reflect variability in numbers of visits and in ancillary and drug costs. They do not include variability in provider time inputs at each visit, covariances between provider time inputs at different visits in an episode, or covariances between time inputs and drug or ancillary costs. These deficiencies are to some extent unavoidable, since our methodology did not entail observation of time inputs at each visit of every episode studied. (The cost of obtaining such observational data would have been considerable.) Consequently, statistical tests relating to our mean episode cost figures must be viewed as suggestive rather than definitive.

Findings

Selected patient and problem characteristics for the 409 otitis media and 390 sore throat episodes, classified by type of provider (physician or nurse practitioner) at the initial visit, are described in Table 3. For nurse practitioner episodes, data are also shown separately for cases with and without physician consultation at the initial visit. In both problem categories, about 10 per cent of initial nurse practitioner visits involved physician consultation. The characteristics of the patients and problems seen by physicians and nurse practitioners appear to be fairly similar. Physician patients with otitis media are slightly younger, but this is not the case for sore throat. The percentages of episodes with multiple problems are about equal, although, as might be expected, nurse practitioner episodes with physician consultation show higher percentages of multiple problems. The relative frequency of serous otitis media is insignificantly higher in nurse practitioner episodes.

Various measures of service utilization for episodes are shown in Table 4. These figures indicate that nurse practitioner patients with otitis media had about twice as many laboratory tests as physician patients, while physician sore throat patients had almost twice as many follow-up visits as nurse practitioner patients. All other differences between physician and nurse practitioner patients are small in relative terms. (This conclusion also holds for x-ray use, which is not shown in Table 4 because it occurred in only seven of the 799 study episodes.)

TABLE 3. Patient and Problem Characteristics for Episodes

	Otitis Media				Sore Throat*			
	MD	NP	NP Only	NP With MD	MD	NP	NP Only	NP With MD
Number of episodes	176	233	204	29	185	205	184	21
Mean patient age (years)	4.85	5.55*	5.56	5.48	7.53	7.52	7.52	7.57
Per cent of patients under 3 years	33.5	19.7*	20.6	13.8	11.4	10.7	10.3	14.3
Per cent with secondary Dx	31.3	39.9	37.3	58.6	25.4	22.4	20.1	42.9
Per cent with serous otitis media	15.3	21.5	21.6	20.7	—	—	—	—

* MD–NP difference is significant at the 0.05 level (two-tailed).

TABLE 4. Utilization Statistics for Episodes

	Otitis Media				Sore Throat			
	MD	NP	NP Only	NP With MD	MD	NP	NP Only	NP With MD
Per cent with follow-up visits	28.4	33.9	33.3	37.9	12.4	6.8*	6.0	14.3
Average number of follow-up visits	0.59	0.57	0.56	0.66	0.15	0.08	0.08	0.14
Average number of MD follow-ups†	0.41	0.26	0.25	0.35	0.12	0.05	0.06	0.05
Average number of NP follow-ups†	0.15	0.25	0.25	0.26	0.03	0.03	0.02	0.10
Average number of NP/MD follow-ups†	0.03	0.06	0.06	0.03	0.00	0.00	0.00	0.00
Per cent with lab tests	17.6	32.2	31.9	34.5	98.4	98.5	98.4	95.2
Average number of lab tests	0.26	0.45*	0.45	0.41	1.42	1.29	1.26	1.52
Per cent with prescriptions	84.7	91.8*	90.7	190.0	59.5	58.5	58.2	61.9

* MD–NP difference is significant at the 0.05 level (two-tailed).

† Significance of MD–NP difference was not computed for this item.

Conversely, differences in mean episode costs, which are reported in Table 5, are fairly large. For otitis media episodes, costs for patients of nurse practitioners are 18 per cent below costs for physician patients; for sore throat episodes, the differential is 25 per cent. Since differences in service utilization are minor, percentage differences in mean ancillary costs are more modest. Nurse practitioner sore throat patients' ancillary costs are 9 per cent lower than those for physician patients, while nurse practitioner otitis media patients' ancillary costs are 14 per cent higher. Thus, the major source of the overall savings for nurse practitioner patients is in the provider-time-related costs. For both problem categories, nurse practitioner provider-time-related costs are about 40 per cent lower.

It has already been noted that the visit time estimates used in the cost calculations were based on small numbers of observed visits, but it is doubtful that sampling errors in these estimates greatly affected the qualitative nature of our findings. While the

TABLE 5. Episode Costs

	Otitis Media				Sore Throat			
	MD	NP	NP Only	NP With MD	MD	NP	NP Only	NP With MD
Mean provider time Related costs ($)	11.05	6.79*	6.24	10.66	7.75	4.59*	4.18	8.21
Mean ancillary costs ($)	7.17	8.19	7.95	9.87	7.89	7.21	7.15	7.71
Mean total costs ($)	18.22	14.98*	14.19	20.53	15.64	11.80*	11.33	15.92

* MD–NP difference is significant at the 0.05 level (two-tailed).

follow-up visit time estimates are especially tentative, the frequency of follow-up visits in the study episodes is low. Moreover, because the cost differences in Table 5 are large, errors in the time estimates would have to be relatively large to explain these differences.

For example, based on the data in Tables 1, 3 and 4, the average physician contact time for a physician otitis media episode was 12.513 minutes (9.3 minutes + 0.41 × 7.5 minutes + 0.03 × 4.6 minutes), while physician contact time for a nurse practitioner otitis media episode was 2.799 minutes (0.26 × 7.5 minutes + [29/233] × 4.6 minutes + 0.06 × 4.6 minutes). To eliminate the mean episode cost differential of $3.24 ($18.22 – $14.98), one would have to reduce the difference in these times by 3.81 minutes, that is, from 9.714 minutes (12.513 minutes – 2.799 minutes) to less than 5.904 minutes.§ This implies that our estimates of physician contact time inputs would have to be at least 64.5 per cent ($\frac{3.81}{5.904} \times 100$) above the true time input figures to explain fully the $3.24 cost differential reported in Table 5. Similar calculations for sore throat episodes show that our estimates of physician contact time inputs would have to be at least 182.2 per cent above the true figures to explain fully the reported cost differential of $3.84. Using the same approach, it can be shown that fully explaining the mean episode cost differentials by underestimates of nurse practitioner contact times would require that the true nurse practitioner time inputs be roughly twice as large as our estimates.

Before concluding that the results presented in Table 5 represent efficiency differences, two other possible explanations should be noted. First, since patients were not randomly assigned to physicians or nurse practitioners, it is possible that physician patients tended to have severe problems. Second, the cost savings achieved by nurse practitioners could simply reflect the provision of lower quality or less effective care.

Tabulation of data from the Columbia EMCRO project (Table 6) does not support either of these explanations. Patient-reported problem severity (i.e., discomfort and activity limitation) at the initial visit did not differ significantly by provider type. Similarly, no significant difference was found in the change in problem status one week after the initial visit. Moreover, while the number of visits included in these tabulations is small, the results in Table 6 are supported by other evidence. In particular, the hypothesis of differences in problem severity by provider type has been rejected in an earlier analysis[22] of Columbia EMCRO data for all types of problems combined. Our own data on patient and problem characteristics (Table 3) also support this conclusion. The hypothesis of provider-type difference in quality or effectiveness were tested in the earlier EMCRO analysis,[22] with data on change in problem status (over one week and one month) and patient satisfaction for all problem types combined, and was rejected.

Summary and Discussion

Concerns with ambulatory health care costs, and with the potential for reducing these costs and improving productivity through the use of NHPs, have been widely discussed. Most previous assessments of this potential have used the visit as the unit of comparison. Important limitations of this method include its failure to account for variations in the use of ancillary and related services, and variations in practice patterns regarding referral and follow-up visits. (Note that in the current study significant physician–nurse practitioner differences were found in the use

---

§ The 3.81 figure is obtained by dividing $3.24 by $0.85, which is the multiplication factor for T_{MD} in our cost formula.

TABLE 6. Initial Problem Status and Change in Problem Status in Otitis Media and Sore Throat Episodes

	Otitis Media				Sore Throat[*]	
	MD	NP[†]	NP Only	NP[‡] With MD	MD	NP Only
Per cent of episodes with:						
No discomfort	16.4	11.8	9.1	28.6	15.4	—
Some discomfort	29.1	21.6	22.7	14.3	19.2	31.4
Moderate discomfort	38.2	45.1	43.2	57.1	46.2	51.4
Severe discomfort	16.4	21.6	25.0	—	19.2	17.1
(N)	(55)	(51)	(44)	(7)	(26)	(35)
Per cent of episodes with:						
No activity limitation	37.0	42.0	41.9	42.9	30.8	19.4
Some activity limitation	31.5	26.0	25.6	28.6	34.6	36.1
Moderate activity limitation	27.8	28.0	27.9	28.6	30.8	44.4
Severe activity limitation	3.7	4.0	4.7	—	3.8	—
(N)	(54)	(50)	(43)	(7)	(26)	(36)
Per cent of episodes with problem status:						
Resolved	27.8	53.8	53.8	—	42.9	75.0
Better	61.1	46.2	46.2	—	35.7	25.0
Unchanged	11.1	—	—	—	21.4	—
(N)	(18)	(13)	(13)	—	(14)	(38)

[*] No NP with MD initial visits for sore throat were reported.

[†] None of the MD–NP differences shown here are significant at the 0.05 level.

[‡] No follow-up data or change in problem status were reported for these visits.

of laboratory tests and drugs for otitis media patients and in follow-up visit rates for sore throat patients.) The comparison of outcome or effectiveness measures for single visits is also problematic. A methodology using the person-year of care overcomes these limitations, but because it involves examination of all services received over a defined period, it is subject to other confounding influences (e.g., random variation in the incidence of health problems, productivity and quality variations among *all* provider organizations serving the study population).

The conceptual advantages of analyses based on episodes of care have been noted by several researchers. In this study, a practical methodology for episode-based comparisons was developed and employed to assess the efficiency of physician and nurse practitioner care for two common pediatric conditions, otitis media and sore throat. (According to our observational study, these two conditions accounted for more than 25 per cent of all pediatric visits.) The results indicated that nurse practitioner care was less costly and that this cost differential could not be attributed to physician–nurse practitioner differences in patient characteristics, problem severity, or effectiveness or quality of care. The reliability of this result is, of course, dependent upon the limitations of our data and methods noted previously. The potential for generalizing this result is also limited to settings with similar organizational and cost characteristics. The methodology, however, can be generally applied. It makes use of existing financial and ambulatory care management information systems, supplemented by a small-scale observational study. The most complex element is the development and application of decision rules to link individual visits and ancillary services systematically into episodes of care. However, previous research[23] has demonstrated that such criteria can be developed empirically and/or by a consensus process among health care providers.

The utility of episode-based cost comparisons (such as that reported in this article) can be seen at several levels. At the clinical level, triage decisions regarding the type of provider to be seen for a problem can be influenced by data such as those presented here. In addition to examining provider differences, this methodology, linked with quality measurements, can be applied to comparisons of different clinical approaches to treatment and follow-up, as well as to the influence of patient education efforts. At the organizational level, productivity information is influential in staffing and salary decisions, and in decisions regarding facilities and the organization of clinical units. On the broader social level, these data help to focus attention on questions of efficiency and effectiveness when care is provided under different modes of organization and financing, and to different population groups.

References

1. Record JC, et al. Cost effectiveness of physician's associates: Kaiser-Permanente experience. American Economic Association Annual Meeting, Dallas, Texas, 1975.

2. Smith KR, et al. An analysis of the optimal use of inputs in the production of medical services. J Hum Resour 1972;7:208.

3. Hoekelman RA. What constitutes adequate well-baby care? Pediatrics 1975;53:313.

4. Schiff DW, et al. The pediatric nurse practitioner in the offices of pediatricians in private practice. Pediatrics 1969;44:62.

5. Charney E, Kitzman H. The child-health nurse (pediatric nurse practitioner) in private practice: A controlled trial. N Engl J Med 1971;285:1353.

6. Denari GA, Katz HP. Periodic quality assessment of a revised well-baby care program in a prepaid multispecialty group practice. Ambulatory Pediatrics Association 14th Annual Meeting, Washington, D.C., 1974.

7. Kane RL, et al. Effects of adding a medex on practice costs and productivity. J Community Health 1978;3:216.

8. Romm J, et al. Survey and evaluation of the physician extender reimbursement study: Cost effectiveness. American Public Health Association Annual Meeting, Los Angeles, 1978.

9. Morris SB, Smith DB. An evaluation of the potential impact of physician extenders on primary care practices. Second International Congress of World Federation of Public Health Association, Halifax, Nova Scotia, 1978.

10. Steinwachs DM, et al. The role of new health practitioners in a prepaid group practice: Changes in the distribution of ambulatory care between physician and nonphysician providers of care. Med Care 1976;14:95.

11. Chambers LW, et al. A controlled trial of the impact of the family practice nurse on volume, quality, and cost of rural health services. Med Care 1977;15:971.

12. Spector R, et al. Medical care by nurses in an internal medicine clinic: Analysis of quality and its costs. JAMA 1975;232:1234.

13. Lewis CE, Resnick B. Nurse clinics and progressive ambulatory patient care. N Engl J Med 1967;277:1236.

14. Scitovsky AA. An index of the cost of medical care—a proposed new approach. The economics of health and medical care. Proceedings of the Conference on the Economics of Health and Medical Care. Ann Arbor: University of Michigan, 1964.

15. Record JC. An episode approach to utilization, costs and effectiveness in health care. Unpublished paper, Kaiser Foundation Health Services Research Center, Portland, 1978.

16. Moscovice I. A framework for analyzing resource use in ambulatory care settings. Med Care 1977;15:1024.

17. Tompkins RK, et al. The effectiveness and cost of acute respiratory illness medical care provided by physicians and algorithm-assisted physician's assistants. Med Care 1977;15:991.

18. U.S. Congress, Congressional Budget Office. Physician Extenders: Their Current and Future Role in Medical Care Delivery. Washington, D.C.: U.S. Government Printing Office, 1979.

19. West SK, et al. Drug utilization review in a health maintenance organization. The Johns Hopkins Medical Institutions. Final report submitted to the National Center for Health Services Research, DHEW, 1977.

20. Skinner EA, et al. Productivity analyses of ambulatory care providers: A method based on observational and information system data. Proceedings of the Annual Conference of the Society for Advanced Medical Systems, Washington, D.C., 1980.

21. Mushlin AI, Appel FA. Testing an outcome-based quality assurance strategy in primary care. Med Care 1980;18(suppl).

22. Levine DM, et al. The role of new health practitioners in a prepaid group practice: Provider differences in process and outcomes of medical care. Med Care 1976;14:326.

23. Steinwachs DM, Yaffe R. An approach to the review of utilization and quality of ambulatory care. Joint Meeting of the Operations Research Society of America and The Institute of Management Science, San Juan, Puerto Rico, 1974.

MEDICAL CARE
February 1982, Vol. XX, No. 2

Costs and Outcomes of Medical Intensive Care

MARK R. CHASSIN, M.D., M.P.P., M.P.H.

Survival, functional status and total six-month medical care costs were measured for 489 consecutive medical intensive care patients. In addition, the study sought to determine if it was possible from data available on admission to identify subgroups of patients with extremely low survival rates. Of the study patients, 86 per cent survived the initial admission, and 72 per cent survived six months. Daily intensive care unit (ICU) charges averaged $1,120 per patient, while total six-month charges per patient averaged $14,577. Age, initial systolic blood pressure, preadmission functional status and diagnosis were all significantly and independently associated with survival. The subgroup with the lowest six-month survival rate (14 per cent) comprised cancer patients with poor preadmission functional status. This study did not identify a subgroup of patients with a survival rate low enough to justify systematic exclusion from intensive care. Controlling ICU costs by excluding patients with very low expected survival may not be possible.

INTENSIVE CARE has spread widely in the last 20 years, yet little is known about its overall effectiveness. The percentage of nonfederal, short-stay hospitals reporting any facilities for intensive care has increased from less than 10 per cent in 1958 to more than 74 per cent in 1977 (Fig. 1).[1,2] While intensive care apparently contributes dramatically to saving individual lives that would previously have been lost, it also generates levels of cost not previously attained. Moreover, this technology contributes to the increasing cost of hospital care. One conservative estimate suggests that intensive care accounts for almost 20 per cent of total hospital costs.[3]

Considering the level of investment in intensive care and the speed with which it has been rising, it is important to define carefully which patients most benefit from it, the degree to which they benefit and the cost of these additional benefits. In 1973, Griner called attention to the lack of research evaluating intensive care.[4] Six years later, Ebert justifiably asserted that intensive care units (ICUs) have proliferated despite a lack of clearly defined indications for patient admissions.[5] This paucity of evaluative research is even more conspicuous in light of mounting pressures to contain the growth of hospital costs.

Of all the varieties of intensive care, medical intensive care has been the least studied. Much work has been done on coronary intensive care,[6–11] and there are

---

From the Department of Medicine, Center for the Health Sciences, University of California at Los Angeles, and The Robert Wood Johnson Foundation Clinical Scholars Program, University of California at Los Angeles.

Supported by a grant from the Robert Wood Johnson Foundation, Princeton, New Jersey. The opinions, conclusions, and proposals in the text are those of the author and do not necessarily represent the views of The Robert Wood Johnson Foundation.

Requests for reprints: Mark R. Chassin, M.D., The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406.

0025-7079/82/0200/0165/$01.25 © J. B. Lippincott Co.

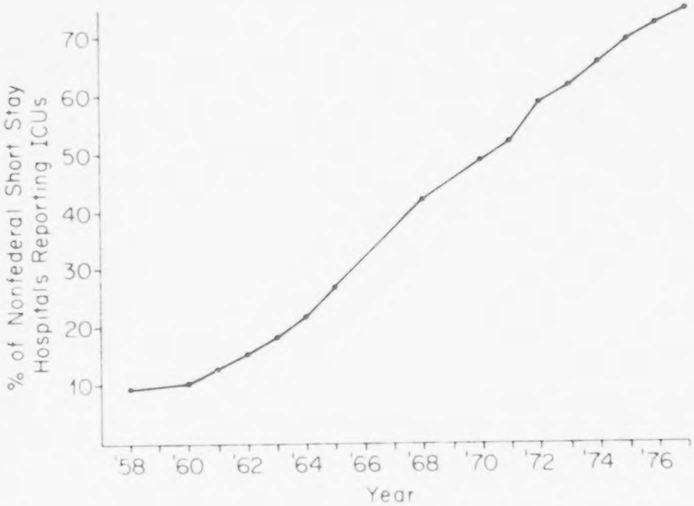


FIG. 1. Rise in percentage of nonfederal, short-stay hospitals reporting availability of intensive care units.
Sources: 1971–77 data[1]
1958–70 data[2]

some published evaluations of surgical intensive care.[12–16] More recently, neonatal intensive care has undergone extensive scrutiny; a randomized clinical trial has been conducted,[17] and the entire experience has been reviewed.[18]

In contrast, literature on the effectiveness and costs of medical intensive care is scarce. Several early studies provided the diagnostic categorization of medical intensive care patients.[19–21] Spagnolo and his coworkers reported a comprehensive set of in-hospital mortality statistics for a general medical intensive care unit (MICU).[22] But only three studies have been reported in which comparisons were made between patients treated in an MICU and comparable patients treated in general ward settings; the first study showed no effect on the survival of patients with pulmonary edema after the opening of an MICU,[23] and the second showed decreased mortality among patients receiving mechanically assisted ventilation following the opening of a respiratory intensive care unit (RICU).[24] The validity of the latter data is, however, open to question because it is not clear that the indications for mechanical ventilatory assistance remained the same in both periods under study. The third study showed no effect on mortality among patients treated for drug overdose after the opening of an RICU, although the rate of complications decreased.[25] The remainder of the published reports give either in-hospital mortality data only or follow-up survival data on a limited range of patients.[26–35] Only one study has reported even limited follow-up data on a general medical intensive care population.[36] Even less information has been published on the costs of medical intensive care,[23,25,35–37] though recently some attention has been given to high-cost patients.[38,39]

In considering how to moderate ICU costs, three different approaches are conceptually apparent. One might try to limit the number of admissions to ICUs by using criteria that more strictly define the appropriate use of intensive care. Second, one might attempt to reduce the length of stay by discharging patients sooner. Finally, one might try to reduce the average cost of a day in an ICU. The present study addresses the feasibility of one variation on the first theme.

One approach to ICU cost control that has received some attention in the literature is to develop criteria for ICU admission and discharge that would exclude from ICUs those patients classified as "hopelessly ill" and for whom "selective limitation of therapeutic measures" is planned[40] or those patients whose short-term prognosis is poor and for whom only "ordinary therapeutic efforts" are planned.[35] Criteria developed so far have been entirely subjective and rely on the attending physician's clinical judgment. There has been little attempt to be explicit about what is meant by "poor short-term prognosis" or about what types of patients might fall into this category.

The study of 489 consecutive medical intensive care patients described here was designed to define one hospital's medical intensive care patients demographically and clinically, and to measure their survival, their functional status and the costs of their care. In addition, the study sought to test the hypothesis that at the time of ICU admission, data are available from the medical record that would allow one to identify subgroups of patients with an especially low likelihood of survival.

Methods

All admissions to the RICU and MICU at a major university teaching hospital in Los Angeles were reviewed for the study period, which extended from January 1, 1977 through July 31, 1977. During this period, 566 patients accounted for 671 admissions to these two ICUs. Ninety-eight per cent of the inpatient medical records were obtained. Of these, 11.6 per cent (64) were rejected from the study for one of four reasons: the patient was a surgical patient, usually admitted for treatment of a post-operative complication (8.0 per cent); the patient was a Coronary Care Unit patient, being attended by CCU staff (2.5 per cent); the patient was transferred to another hospital before leaving the ICU (0.7 per cent); or the patient left the ICU against medical advice (0.4 per cent). The remaining 489 patients form the sample that was studied further.

The RICU at the study hospital comprises eight beds and is staffed by a team of physicians including house staff, a pulmonary fellow and one attending physician from the pulmonary division. The team is separate from the general medicine ward teams. All admissions to the RICU were approved by the pulmonary fellow during the time of the study. The MICU also consists of eight beds. Patients in this unit were managed by the general medicine ward teams during the study period. Admissions to this unit were usually arranged by the senior admitting resident in the emergency department. Both ICUs had similar nurse staffing patterns, including the capacity for one-to-one nursing when indicated, and both had access to similar equipment. In the event that both the RICU and the MICU were full, a medical intensive care patient might be admitted to a nonmedical ICU. Some of these patients would be transferred to either the RICU or the MICU, if medically appropriate, when a bed became available. Such patients are included in the present study. Patients receiving all of their medical intensive care in nonmedical ICUs are, however, not included. The term "medical ICUs" will be used in the remainder of this report to include both the RICU and the MICU of the study hospital.

Each hospital record was reviewed for demographic and diagnostic data, as well as for clinical and laboratory data relevant to severity of illness. In patients with multiple diagnoses, it was uniformly possible to select the one major disease process that was most life-threatening and for which the ICU admission was required. In addition, medical and surgical procedures and physician visits were tabulated by type of service and according to whether the patient was in the ICU or on a ward when the service was rendered.

Each patient's hospital bill was obtained and subdivided by kind of charge and by the location of the patient at the time the charge was incurred. A list of medical procedure fees was obtained from the hospital division that prepares the bills for such procedures. The following pricing decisions for charges for physician inpatient services were made to reflect an average or low-average level of charge for the indicated service within the geographic area served by the study hospital. Each admission history and physical examination by an attending physician, and each new consultation was valued at $120; each follow-up visit by an attending physician or consultant, as documented by a signed note in the medical record, was valued at $30. Surgical fees were determined by reference to the 1964 California Relative Value Studies. Each unit was assigned a weight of $55.

Next, each patient or surviving family member was interviewed by telephone. The interviews were conducted from six to 20 months following discharge (median 12 months). Data were obtained at that time concerning patients' functional status six months after discharge and their functional status during the time preceding the initial hospitalization. The questions used to assess functional status were adapted from the physical limitations scale used in the Rand Health Insurance Study.[41] In addition, patients were asked about their utilization of medical services during the six months after discharge; data were collected on the frequency of physician visits, laboratory and x-ray examinations, other tests and procedures, drugs and equipment, treatments by nonphysician medical personnel, rehospitalization and residence in an extended care facility.

Finally, when patients or family members were unable to provide cost data for these follow-up services, as was most often the case, each local physician, hospital and nursing home was contacted and the charges for their services recorded. A survey was conducted of three pharmacies in the area to price the drug regimens followed by study patients. The prices charged by the study hospital's pharmacy, a private hospital pharmacy, and a popular discount pharmacy were averaged to arrive at a single price for each drug used by these patients. The regimen described by the patient in the interview was then priced accordingly.

No distinction is drawn in this report between charge and cost because of the difficulty in obtaining data on the true costs of medical services. In most cases charge data are adequate estimates of true costs. It is important to realize, however, that in certain areas, particularly inpatient ancillary services, charges may exceed true costs.

Of the 489 initial hospital bills, all but six were obtained for analysis. During the six months following discharge, the survivors had a total of 344 readmissions to hospital. Only four of the 207 bills for readmission to the study hospital were not obtained, and only nine of the 137 bills from other hospitals were not available. No difficulty was encountered in obtaining charge data from area physicians or nursing homes. Survival data were obtained on 97 per cent of the sample. Functional status data were not obtained for 39 patients (8 per cent). Ten patients (2 per cent) refused to be interviewed. In 29 cases (6 per cent) neither the patient nor any close family member or friend was available for the interview, but survival data were obtained for 12 of these cases.

Results

A histogram of the group's age distribution is presented in Figure 2; the mean age was 54 years. Fifty-four per cent of the patients were female. Table 1 presents the frequency of the major diagnostic categories responsible for the medical ICU admissions of these patients and the most common subgroups within the first five

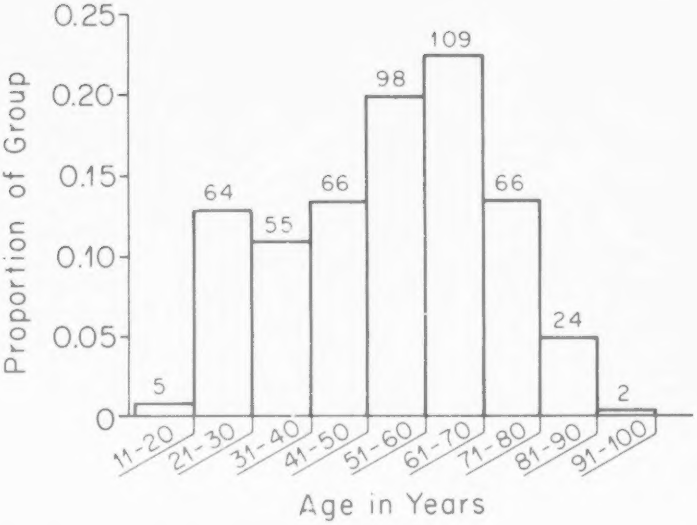


FIG. 2. Histogram of age distribution of medical intensive care population. Numbers above histogram bars refer to numbers of patients in each age group.

categories. Each category represents a broad range of diseases, and only three homogeneous diagnostic groups each contain 5 per cent or more of the sample: chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF) and overdose. These medical ICUs, therefore, treat a wide spectrum of illnesses, each of which represents a very small proportion of the total.

The average length of ICU stay was 5.1 days, with a range of 1 to 95 days. Patients who stayed only one day in the ICU composed 27 per cent of the sample, patients who stayed two days made up an additional 22 per cent, while another 10 per cent stayed more than ten days. The total hospital stay was similarly widely distributed, with a mean of 13.5 days and a range of 1 to 120 days.

Of the 489 study patients, 70 died during the initial hospital admission, accounting for an in-hospital survival rate of 86 per cent. At six months following discharge, 134 patients had died and 17 had been lost to follow-up, accounting for a survival rate of 72 per cent at six months. Figure 3, showing the survival curve for the group as a whole using the life table method of analysis, reveals a one-year survival rate of 66 per cent.

The functional status questionnaire permitted division of the group into six

TABLE 1. Major Diagnostic Categories and Selected Subgroups

Category	Number of Patients		Per Cent	
Cardiovascular	110		22.5	
Congestive heart failure		38		7.8
Hypertension		16		3.3
Pulmonary	106		21.7	
COPD		48		9.8
Pneumonia		14		2.9
Gastrointestinal	73		14.9	
Duodenal ulcer		15		3.1
Gastritis		13		2.7
Inflammatory bowel disease		10		2.0
Cancer	59		12.1	
Solid tumors		37		7.6
Hematologic		22		4.5
Neurologic	36		7.4	
CVA (cortical)		14		2.9
Overdose	28		5.7	
Renal	20		4.1	
Rheumatologic	19		3.9	
Diabetes mellitus	15		3.1	
Other	23		4.7	
Total	489		100.1	

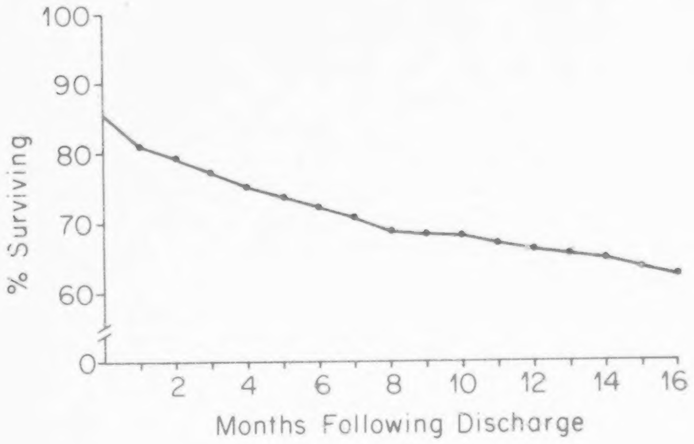


FIG. 3. Survival curve of medical intensive care patient population.

This curve was constructed using the life table method.

classes based on the ability to perform certain activities. Table 2 presents the distribution of functional status classes before admission and at six months after discharge. Considering only survivors, the two distributions are virtually identical. Of the 450 patients for whom functional status data were available, 42 per cent reported that they were in the same functional class six months after discharge as they had been before hospitalization. Only 8 per cent reported an improved functional status, while 20 per cent reported worsened functional capacities. The remaining 30 per cent were dead.

TABLE 2. Distribution of Functional Status

		Six Months After Discharge	
Class*	Before Hospitalization (%)	Whole Group (%)	Survivors (%)
5	29	18	26
4	13	11	16
3	21	13	18
2	21	15	22
1	6	5	7
0	10	8	11
dead	—	30	—

* Definitions of functional classes:

Class 5: Able to perform vigorous activities such as running or lifting heavy objects.
Class 4: Unable to perform vigorous activities, but able to walk several blocks.
Class 3: Unable to walk several blocks, but able to walk one block.
Class 2: Able to walk alone, but not as far as one block.
Class 1: Unable to walk alone, but able to eat, dress, and bathe without assistance.
Class 0: Unable to walk or perform self-care activities without assistance.

Survival

Table 3 presents the univariate relationships between survival and each of many patient characteristics for which statistically significant associations were demonstrated. The chi-square test was used to evaluate the statistical significance of these associations. Diagnosis was the most effective characteristic for distinguishing those subgroups of patients with very high survival rates from those with very low ones. Patients with cancer, including both solid tumors and hematologic malignancies, who were sick enough to be treated in these medical ICUs fared the worst. No other way of categorizing the patient population produced a group with a lower six-month survival rate than the 24 per cent experienced by the cancer patients. Nor did any other method of classification produce a group with a higher rate than the patients with hypertension, 93 per cent of whom survived six months.

TABLE 3. Univariate Associations with Survival

Characteristic	Number of Patients	Six-month Survival (%)
TOTAL GROUP	472	71.6
Sex		
Male	213	66.7
Female	259	75.7*
Age		
30 and under	65	89.2
31–50	112	76.8
51–70	204	66.2
over 70	91	64.8†
Initial systolic blood pressure		
100 mmHg and less	65	58.5
over 100 mmHg	407	73.7‡
Preadmission functional status		
Class 0	45	44.4
Class 1	27	59.3
Class 2	96	63.5
Class 3	94	70.2
Class 4	59	81.4
Class 5	129	81.4†
Preadmission employment status		
Full-time	105	83.8
Part-time	29	58.6
Housewife	43	81.4
Student	20	85.0
Not working	253	62.8†
Initial mental status (excluding overdoses)		
Alert	343	77.3
Lethargic	50	58.0
Stuporous	31	38.7
Semi-comatose and comatose	20	30.0†
Diagnosis		
Hypertension	14	92.9
Overdose	27	92.6
Diabetic ketoacidosis	10	90.0
COPD	47	85.1
Upper gastrointestinal bleeding	60	80.0
Pneumonia	14	78.6
Stroke (cortical)	14	71.4
Chronic renal failure	16	62.5
Congestive heart failure	37	59.5
Cancer	59	23.7†

* = $p < 0.05$.
† = $p < 0.005$.
‡ = $p < 0.025$.

Multivariate analysis confirmed and elucidated further the relationships described in Table 3. The patient population was divided into eight subgroups on the basis of age, initial systolic blood pressure and preadmission functional status; each variable was dichotomized. These particular variables were chosen for further analysis because of their strong univariate associations with survival. For each variable, the dividing line was chosen so as to maximize the difference between the two subgroups so created. Comparison of the six-month survival of these subgroups revealed a uniform trend. Patients who were older, who were admitted to the ICU in shock or who had low preadmission functional status did less well than their respective counterparts. This was established by comparing the effect of each variable on survival while controlling for the other two. The subgroup with the best six-month survival—those under 50 years of age, admitted with systolic blood pressures above 100 mmHg, and with preadmission functional status of class 4 or 5—experienced a survival rate of 90 per cent. In the subgroup with the lowest survival—those over 50, with admitting systolic blood pressures of 100 mmHg or less, and in the lowest four functional classes prior to admission—only 39 per cent were alive six months after discharge. Even this subgroup of 18 patients, however, had a better survival experience than did the patients with cancer.

The independent association of these three variables with survival is noted also when one looks within diagnostic groups. Thus, for patients with congestive heart failure, those who were 60 years of age or younger experienced a six-month survival rate of 85 per cent, while those older than 60 experienced a rate of only 50 per cent. Similarly, those CHF patients who reported preadmission functional status in classes 3–5 survived at a rate of 71 per cent, while those in the lower functional classes survived at a rate of 47 per cent. Because of small sample sizes, neither of these large differences is statistically significant.

For patients admitted to these medical ICUs with upper gastrointestinal bleeding, 84 per cent of those 60 years old or younger survived six months, while 73 per cent of those over 60 survived. Of those admitted with systolic blood pressures initially over 100 mmHg, 82 per cent lived through the follow-up period, but only 71 per cent of those admitted with initial systolic pressures of 100 mmHg or less were alive six months after discharge. Further, 88 per cent of those in the highest two functional classes survived, while only 69 per cent of those in functional classes 0–3 lived. None of these differences is statistically significant.

Finally, for patients with cancer, those in functional classes 4–5 had a six-month survival rate of 43 per cent, but those in classes 0–3 experienced a rate of only 14 per cent; this difference is statistically significant at the 5 per cent level. The subgroup of 36 cancer patients with poor preadmission functional status experienced the lowest survival rate of any subgroup identified in the study.

Costs

The average daily ICU charge for this patient population exceeds $1,000 per day, as seen in Table 4. It can readily be appreciated from these data that the largest single component of the ICU segment of the hospital bill is represented by hospital services, which constitute 44 per cent of the bill; laboratory charges account for 23 per cent, and professional fees make up 14 per cent. If one adds the charges for hospital care in the follow-up period ($3,900) to those for the initial hospital stay ($9,601), it is clear that hospital care accounts for 93 per cent of average total six-month charges.

There was little variation in these average charges by age, initial systolic blood pressure, preadmission functional status or initial mental status. There was significant variation by diagnosis; diabetic ketoacidosis had the lowest six-month average ($9,899) and chronic renal failure the highest, with a value of $31,668. Survivors averaged $12,860, while nonsurvivors averaged $19,838 for the six months. Both of these differences are statistically significant ($p < 0.005$, Student's t-test). Interestingly, the average charges accumulated by those nonsurvivors who died during the initial hospitalization ($19,812) were virtually identical to those incurred by the nonsurvivors who died during the follow-up period ($19,868).

Figure 4 shows the extent to which resources for intensive care patients are spread unevenly across the patient population; 17 per cent of the patients consumed half of the total resources. A subgroup of 32 patients (7.4 per cent), each of whom incurred more than $40,000 in total six-month charges, was designated as the high-cost subgroup and examined more closely. This subgroup accounted for 31 per cent of the total resources. The average charge in this subgroup was $62,736, and the range was from $40,432 to $135,547.

Table 5 compares selected characteristics of the high-cost subgroup with those of the group as a whole. Mean age, functional status distribution and proportion of severely hypotensive patients are similar between the two groups. There are slightly more males in the high-cost subgroup and more patients admitted with an abnormal mental status. As expected, the ICU length of stay for these patients was markedly longer than that for the entire group. There were relatively fewer patients with cancer in the high-cost subgroup, but relatively more with chronic renal failure and rheumatologic conditions. Survival was modestly, but not dramatically, reduced.

Thus while there are some differences between the high-cost subgroup and the medical intensive care population, in general the high-cost subgroup is broadly representative of the whole. It is not narrowly confined to one age group, functional class or diagnostic category, nor does it have a uniformly poor outcome. The only striking difference, length of stay, was clearly dictated by the selection of the patients with the highest costs. The close correlation between length of stay and cost is clear; the longer one stays in a hospital, on average, the higher the cost will be. This association, however, is of limited usefulness in identifying patients who are likely to use large amounts of resources before those resources are consumed. It was not possible in this study to identify any characteristics available on admission that would identify those patients who would later enter the high-cost subgroup.

TABLE 4. Analysis of Average Total Six-Month Charges

Type of Service	Total ($)	Cost Per Day ($)
Initial Hospitalization		
ICU hospital services*	2,485	489
ICU inhalation therapy	317	62
ICU pharmacy†	452	89
ICU X-ray‡	317	62
ICU laboratory§	1,318	259
ICU professional fees‖	809	159
Average total ICU charges	5,698	1,120
Average total ward charges	3,903	
Average total initial hospitalization	9,601	
Average charges for rehospitalization during six-month follow-up period	3,900	
All other charges during six-month follow-up period	1,076	
Average total follow-up charges¶	4,976	
Average total six-month charges per patient	14,577	

* Includes charges for room and board, hemodialysis, emergency and operating rooms, supplies, and physical therapy.

† Includes charges for all medications and intravenous fluids.

‡ Includes charges for diagnostic radiology, ultrasound (except echocardiography), nuclear medicine and radiation therapy.

§ Includes all clinical laboratory charges as well as charges for electrocardiography, echocardiography, blood gas analyses, outside laboratory tests, and pathology.

‖ Includes charges for visits, consultations and procedures, including those for anesthesia.

¶ Includes nonsurvivors in a denominator.

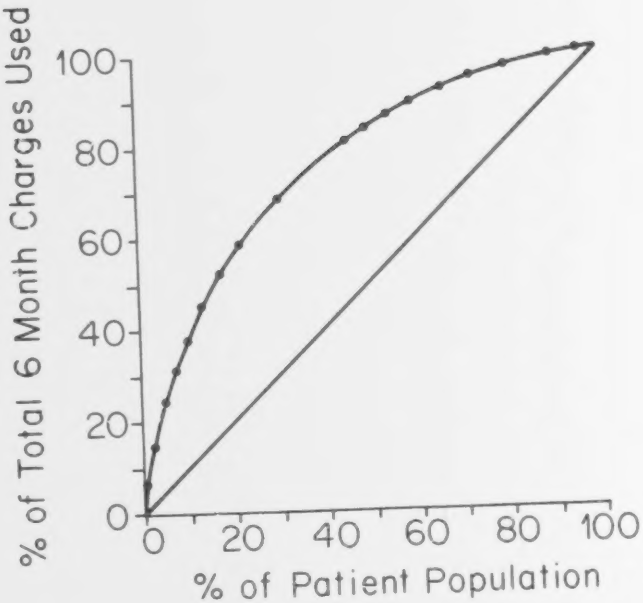


FIG. 4. Distribution of total six-month charges. Curved line represents actual distribution of charges; straight line indicates what distribution would have been had each patient incurred the same charges.

Discussion

Currently, very little is known about what kinds of patients are hospitalized in MICUs, and there is even less knowledge of the extent to which different institutions serve diverse MICU populations. How, for example, do teaching hospitals differ in this regard from community hospitals without house staff? Are there differences related to hospital or unit size? What are the urban/rural differences? Any attempt to develop guidelines for MICU use must be based on a thorough understanding of the differences in the populations served by these units.

One goal of the present study was to characterize a single hospital's medical intensive care population. The patients were found to be relatively young (mean age 54 years, only 19 per cent over 70 years of age). The diagnostic entities represented in the medical ICU population were spread over a wide range; only three homogeneous diagnostic groups each contained more than 5 per cent of the population. If this finding is confirmed in other institutions, the task of defining appropriate guidelines for MICU utilization will be that much more complex.

The present study identified a number of variables that appeared to be independently associated with survival: age, preadmission functional status, initial systolic blood pressure and diagnosis. This study also illustrates one particular problem with performing multivariate analysis on data from intensive care populations. Even with almost 500 study patients already screened to exclude surgical and CCU patients, controlling for more than three variables, particularly if one of them is diagnosis, partitions the group into very small subgroups. Such small sample sizes prevent even large differences in survival from attaining conventional levels of statistical significance. The findings of the study

TABLE 5. Characteristics of the High-Cost Subgroup

Characteristic	High-Cost Subgroup (%)	Whole Group (%)
Sex		
Male	53.1	45.1
Female	46.9	54.9
Mean age	53 years	54 years
Initial systolic blood pressure		
100 mmHg and less	12.5	13.8
Over 100 mmHg	87.5	86.2
Preadmission functional status		
Classes 0–1	12.9	16.0
Classes 2–3	45.2	42.2
Classes 4–5	42.0	41.8
Initial mental status (excluding overdoses)		
Alert	62.5	77.1
Abnormal	37.5	22.9
Length of ICU stay		
1–2 days	9.4	49.6
Over 10 days	56.2	10.3
Diagnosis		
Cardiovascular	25.0	22.5
Pulmonary	15.6	21.7
Gastrointestinal	15.6	14.9
Neurologic	6.2	7.4
Cancer	6.2	12.1
Chronic renal failure	12.5	3.3
Rheumatologic	12.5	3.9
Survival		
In-hospital	71.9	85.7
Six-month	53.1	71.6

with respect to survival should be confirmed in prospective studies at different institutions with larger numbers of patients. These further investigations will determine the extent to which the results of this study are generalizable and will permit a more detailed multivariate analysis of predictors of survival.

The finding that a simple functional status assessment yields important information about survival also requires confirmation in prospective studies. The retrospective use of this instrument in this study revealed that half of the patients perceived their functional status six months after discharge as equal to or better than their preadmission status. The direction of any bias that might have been introduced by the application of the instrument retrospectively is difficult to assess. On the one hand, relatives of nonsurvivors might tend to see the preadmission functional status of their relatives as worse than it actually was, because they remember only the severity of the illness. On the other hand, they might tend to recall the preadmission status as better than it actually was, because they remember the relatives as they were before the terminal illness. Similar arguments apply to patients recalling their own preadmission functional status. Prospective use of the instrument is the only way to measure the significance of these possibly confounding effects. If this measure is validated prospectively, it may prove a useful way to assess intensive care patient outcomes in addition to the more usual but cruder mortality measure.

There are relatively few data with which to compare the charge data gathered in this study. Other studies that have presented information on the costs of care for seriously ill patients analyzed very different kinds of patients in different time periods. After adjusting for inflation, the present investigation found costs that were roughly comparable to those measured by Cullen and his coworkers in their study of the most severely ill of their surgical ICU patients from 1972-73.[15] After a similar inflation adjustment, the present study also found costs roughly comparable to those measured by Griner for patients with pulmonary edema from 1970-71.[23] In their study of medical intensive care, Thibault and his colleagues measured hospital charges that were somewhat lower than those in this study for several diagnostic categories.[36] On the other hand, this study found costs substantially lower than those measured by Turnbull and his colleagues for cancer patients[35] and less than recent estimates of the cost of an institutionally based program of maintenance hemodialysis.[42] Unlike the present study, however, none of the studies previously cited separated ICU charges from ward charges. Comparisons were therefore made of total hospital charges, and no conclusions can be drawn regarding the relative charges generated by these different MICUs.

Until now much of the discussion of controlling ICU costs has focused on excluding patients with poor short-term prognoses from ICUs. A discussion of the limited conceptual and empirical evidence for the effectiveness of utilization review in containing hospital costs in general has been presented elsewhere[43]; it is equally applicable to this form of utilization review for intensive care patients. In the absence of explicit criteria for what is meant by poor short-term prognosis, physicians have been called upon to exercise their clinical judgment in deciding whom to admit to ICUs and whom to exclude. Rodman and his colleagues have shown that experienced critical care physicians were poorly able to distinguish eventual survivors from nonsurvivors in a mixed medical-surgical ICU, even at two and seven days after ICU admission.[44] This finding, if confirmed, suggests that more explicit criteria, based on empirical data, will be necessary if this effort is to be successful.

The present study sought to identify subgroups of medical intensive care patients with extremely low expected survival—categories of patients who might be candidates for exclusion from ICUs on the grounds of having poor short-term prognoses. The subgroup with the lowest six-month survival rate comprised cancer patients whose preadmission functional status placed them in classes 0–3; their 14 per cent survival rate is, however, clearly not low enough to justify systematic exclusion from ICUs.

In this study population, simple rules did not succeed in defining subgroups with extremely low survival rates. One cannot, for example, arbitrarily exclude all those over a certain age or all those with preexisting disabilities. Of the 15 study patients who were both older than 80 years of age and in the three lowest functional classes prior to their intensive care hospitalization, 12 survived to be discharged, and six survived six months.

One possible explanation of this finding is that the study hospital has already succeeded in excluding patients with very low expected survival from its medical ICUs. This was not, however, an explicit hospital policy at the time of the study. Decisions to admit or not to admit patients to intensive care were made by individual physicians and families. While it is likely that some critically ill patients were not admitted to intensive care as a result of this process, this fact seems an unlikely explanation for the inability of this study to define with simple rules subgroups of patients with extremely low rates of survival. Moreover, one would expect that a tertiary hospital would have a relatively sicker intensive care population than other hospitals. Therefore, if one does not find patients with very low survival in a tertiary hospital, it may not be easier to do so in other, less specialized institutions.

Furthermore, each intensive care patient represents a virtually unique combination of primary and secondary disease processes, therapeutic regimens, and complications of therapy and disease. While in some sense this may be true for all patients, the significance of the differences among patients with similar diagnoses seems magnified in an intensive care population. This uniqueness, which is reflected in the diagnostic diversity described previously, makes the task of setting criteria for ICU exclusion even more difficult.

A final problem with putting this kind of control system into practice is the extreme difficulty encountered in obtaining agreement as to what level of expected survival rate would be low enough to justify exclusion from intensive care. Would survival have to be unprecedented, a situation that has been described for some badly burned patients?[45] Is 5 per cent low enough—a 1 in 20 chance? Furthermore, how should the views of patients and families be added into the balance? By now we are familiar with patients and families who decline heroic measures. But, conversely, should a patient's or family's desire for intensive care always supersede our criteria for exclusion, however they have been developed? The imponderability of these questions raises serious doubts about whether the effort to set criteria should be attempted at all.

This study demonstrated that, at least for medical intensive care, no single type of patient was responsible for a large portion of the total costs. Rather, the high-cost subgroup was spread among many ages, diagnostic groups and disability classes. This finding suggests that cost control cannot be achieved by focusing attention on any one homogeneous group of patients. Intensive care as a whole must be scrutinized more carefully.

Furthermore, the fact that more than half of the high-cost subgroup survived six months suggests that this investment of resources was not entirely without benefits. This finding highlights the very difficult choices that face intensive care physicians. Increasingly, we confront clinical situations in which the use of larger and larger amounts of medical resources may be associated with smaller and smaller additional benefits.

Measures that are more morally tenable exist for controlling ICU costs. First, one can examine a particular ICU population to ensure that no one has been admitted who is too healthy for intensive care. In the present study, for example, there was a subgroup of 16 patients among the 60 admitted for upper gastrointestinal bleeding who received no blood transfusions during their hospitalization; none of these patients died in the hospital, and only one died during the six-month follow-up period. Furthermore, of the 16 patients, 11 were admitted to the ICU with normal hematocrits, and seven left the hospital with normal hematocrits. In retrospect, it is reasonable to question their need for intensive care. Further investigation may lead to a clearer understanding of when or whether such patients require intensive care.

The second way in which physicians can influence ICU costs is through a careful attempt to reduce the intensity of intensive care. The present study measured average daily ICU charges of $1,120, a figure that includes $62 per day for inhalation therapy, a charge equal to the average daily charge for all radiology services. Were all of these services necessary and effective? By how much could the $259 per day in laboratory charges be reduced if tests were ordered more prudently? These charges may overstate the opportunity for real cost savings, because charges for ancillary services may exceed their true costs. Substantial research will be required to determine which services are necessary and effective, and under what circumstances. The fruits of this effort, however, can result in better patient care at lower risk and lower cost.

The juxtaposition of highly visible and rapidly growing costs with less apparent and inadequately documented benefits can be expected to continue to attract public attention. It is incumbent upon those clinicians with an interest in effective intensive care to design and execute the research necessary to establish the place of intensive care in modern medicine. If this task is not addressed in a timely fashion, arbitrary restrictions on the use of intensive care issued by third-party payors may ensue.

Acknowledgments

The author is indebted to Linda Akita, R.N., Denice Allen, John Herman, Paul Hoffman, Susan Keel, R.N., J.D., Marion Winston, R.N., and Cheryl McCarthy, R.N., for their considerable assistance in data collection and processing; to Drs. John Beck, Robert Kane, Joseph Newhouse, Helen Smits and David Solomon for their insightful comments on earlier versions of this paper; to Marilyn Dean for her cogent suggestions on the final version; and to Dr. Robert Brook for his perceptive guidance throughout the study.

References

1. American Hospital Association. Hospital Statistics: 1971–1977 editions. Chicago, 1971–1977.

2. American Hospital Association. Hospitals 1959–1971, 33–45, Guide Issues.

3. Russell LB. The role of technology assessment in cost control. Presented at "Critical Issues in Medical Technology," conference sponsored by Harvard-MIT Division of Health Sciences and Technology, Boston, April 1980.

4. Griner P. Medical intensive care in the teaching hospital: Costs versus benefits. The need for an assessment Ann Intern Med 1973;78:581.

5. Ebert RE. Effectiveness and efficiency—How well are we doing? Am J Med 1979;66:191.

6. Martin SP, Donaldson MC, London CD, et al. Inputs into coronary care during 30 years. A cost effectiveness study. Ann Intern Med 1974;81:289.

7. Hofvendahl S. Influence of treatment in a coronary care unit on prognosis in acute myocardial infarction. Acta Med Scand 1973;519(suppl):1.

8. Hubner PJB, Goldberg MJ, Lawson CW. Value of routine cardiac monitoring in the management of acute myocardial infarction outside a coronary care unit. Br Med J 1969;1:815.

9. Reynell PC, Reynell MC. The cost-benefit analysis of a coronary care unit. Br Heart J 1972;34:897.

10. Stross JK, Willis PW III, Reynolds EW Jr. Effectiveness of coronary care units in small community hospitals. Ann Intern Med 1976;85:709.

11. Bloom BS, Peterson OL. End results, cost and productivity of coronary-care units. N Engl J Med 1973;288:72.

12. Pessi TV. Experiences gained in intensive care of surgical patients. A prospective clinical study of 1001 consecutively treated patients in a surgical intensive care unit. Ann Chir Gynaecol 1973; 62(suppl):1.

13. Davidson IA, Bargh W, Cruickshank AN, et al. Crush injuries of the chest. A follow-up study of patients treated in an artificial ventilation unit. Thorax 1969;24:563.

14. Morgan A, Daly C, Murawski BJ. Dollar and human costs of intensive care. J Surg Res 1973;14:441.

15. Cullen DJ, Ferrara LC, Briggs BA, et al. Survival, hospitalization charges and follow-up results in critically ill patients. N Engl J Med 1976;294:982.

16. Franco D, Bismuth H. Five years of experience with an intensive care unit specializing in abdominal surgery. Prog Surg 1977;15:37.

17. Renou P, Chang A, Anderson I, et al. Controlled trial of fetal intensive care. Am J Obstet Gynecol 1976;126:470.

18. Thompson T, Reynolds J. The results of intensive care therapy for neonates. J Perinat Med 1977;5:59.

19. Callahan JA, Spiekerman RE, Broadbent JC, et al. St. Mary's Hospital-Mayo Clinic medical intensive care units. II. Patient population. Mayo Clin Proc 1967;42:332.

20. Lepore MJ, Grace WJ. Role of the intensive care unit in gastroenterology 1969;51:493.

21. Clark TJH, Collins JV, Evans TR, et al. A review of experience operating a general medical intensive care unit. Br Med J 1971;1:158.

22. Spagnolo SV, Hershberg PI, Zimmerman HJ. Medical intensive care unit: Mortality rate experience in large teaching hospital. NY State J Med 1973;73:754.

23. Griner P. Treatment of acute pulmonary edema: Conventional or intensive care? Ann Intern Med 1972;77:501.

24. Rogers RM, Weiler C, Ruppenthal B. Impact of the respiratory intensive care unit on survival of patients with acute respiratory failure. Chest 1972;62:94.

25. Piper KW, Griner PF. Suicide attempts with drug overdose: outcomes of intensive vs. conventional floor care. Arch Intern Med 1974;134:703.

26. Askrog VF, Dam WH, Hagelsten JO. Treatment of patients with pulmonary oedema in an intensive therapy department. Br J Anaesth 1967,39:252.

27. Sluiter HJ, Blokzijl EJ, van Dijl W, et al. Conservative and respirator therapy of acute respiratory insufficiency in patients with chronic obstructive lung disease. Am Rev Respir Dis 1972;105:932.

28. Jessen O, Kristensen HS, Rasmussen K. Tracheostomy and artificial ventilation in chronic lung disease. Lancet 1967;ii:9.

29. Bendixen HH. The cost of intensive care. In: Bunker JP, Barnes BA, Mosteller F, eds. Costs, Risks, and Benefits of Surgery. New York: Oxford University Press, 1977.

30. Linton RC, Walker FW, Spoerel WE. Respirator care in a general hospital: A five-year survey. Can Anaesth Soc J 1965;12:450.

31. Campbell D, Reid JM, Telfer ABM, et al. Four years of respiratory intensive care. Br Med J 1967;4:255.

32. Petty TL, Lakshminarayan S, Sahn SA, et al. Intensive respiratory care unit. Review of ten years' experience. JAMA 1975;233:34.

33. Soler NG, Bennett MA, FitzGerald MG, et al. Intensive care in the management of diabetic ketoacidosis. Lancet 1973;i:951.

34. Nunn JF, Milledge JS, Singaraya J. Survival of patients ventilated in an intensive therapy unit. Br Med J 1979;1:1525.

35. Turnbull AD, Carlon G, Baron R, et al. The inverse relationship between cost and survival in the critically ill cancer patient. Crit Care Med 1979;7:20.

36. Thibault GE, Mulley AG, Barnett GO, et al. Medical intensive care: Indications, interventions, and outcomes. N Engl J Med 1980;302:938.

37. Griner PF, Liptzin B. Use of the laboratory in a teaching hospital. Implications for patient care, education, and hospital costs. Ann Intern Med 1971;75:157.

38. Schroeder SA, Showstack JA, Roberts HE. Frequency and clinical description of high-cost patients in 17 acute-care hospitals. N Engl J Med 1979; 300:1306.

39. Zook CJ, Moore FD. High-cost users of medical care. N Engl J Med 1980;302:996.

40. A report of the Clinical Care Committee of the Massachusetts General Hospital: Optimum care for hopelessly ill patients. N Engl J Med 1976;295:362.

41. Stewart AL, Ware JE Jr, Brook RH, et al. Conceptualization and Measurement of Health for Adults in the Health Insurance Study. Volume II, Physical Health in Terms of Functioning. R-1987/2-HEW. Santa Monica: The Rand Corporation, July 1978.

42. Stange PV, Sumner AT. Predicting treatment costs and life expectancy for end-stage renal disease. N Engl J Med 1978;298:372.

43. Chassin MR. The containment of hospital costs: A strategic assessment. Med Care 1978;16(suppl): October: 1.

44. Rodman GH, Etling T, Civetta JM, et al. How accurate is clinical judgment? Crit Care Med 1978;6:127 (abstract).

45. Imbus SH, Zawacki BE. Autonomy for burned patients when survival is unprecedented. N Engl J Med 1977;297:308.

MEDICAL CARE
February 1982, Vol. XX, No. 2

Predicting the Outcome of Primary Care

DIANA DRYER WRIGHT, M.D.,* AND ROBERT L. KANE, M.D.†

Each physician's ability to treat disease is limited by his/her ability to discriminate among patients on the basis of risk. The relationships of physician expectation of outcome (prognosis) to actual outcome and to cost were determined for 1,757 patients seeking primary care. Outcome was measured by a seven-level functional-status scale; patients who returned to their usual level of function after an acute illness were defined as having good outcomes. Although 24 per cent of patients experienced bad outcomes, physicians had anticipated only 6 per cent. Physician's predictions of bad outcomes had a sensitivity of 13.6 per cent and a specificity of 96.9 per cent. Patients with bad outcomes had slightly higher laboratory costs than did patients with good outcomes, but a much larger increase was seen in laboratory, office and total costs for all patients for whom physicians expected bad outcomes, regardless of the actual results. A feedback loop is recommended to provide a better perspective on outcome and eventually to improve the efficiency and cost benefit of the medical decision-making process.

THE INCREASING COST of medical care has made primary-care physicians aware that, as an integral part of any health care system, they must increase their efforts for more efficient, as well as more effective, treatment.[1,2] Both their patients and those responsible for third-party payment are demanding answers about what they are buying and how much they should be paying for it.[3] Physicians are exploring how current methods of assessing quality of care are related to cost.[4,5]

Primary-care physicians face a special challenge because they have been identified as the entry point into the health care system. In a context of high volume, the physician must discriminate between substantive and self-limited illness in order to affect outcome.[6] Traditional clinical judgment has been shown to be relatively imprecise, even in such defined clinical entities as systemic lupus erythematosis[7] and peptic ulcer disease.[8] Increased reliance on laboratory tests is not necessarily related to improvement in outcome.[9-12]

Far less information is available about the course of the cluster of complaints that represent the bulk of primary care; the typical diagnosis-oriented approach may not be useful in evaluating the course of mild, nonspecific or self-limited illnesses. The assumption is that a diagnosis implies a prognosis,[13] but too often diagnostic deci-

---

* Department of Family and Community Medicine, University of Utah College of Medicine.

† School of Medicine, UCLA, and The Rand Corporation, Santa Monica, California.

Work for this study was supported by Public Health Service Award No. 1 RO1 HSO 1596 from the National Center for Health Services Research.

Reprint requests: Robert L. Kane, M.D., The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406.

0025-7079/82/0200/0180/$00.90 © J. B. Lippincott Co.

sions must be made on the basis of inadequate data by physicians who have not been trained to deal with uncertainty. The primary-care physician's ability to predict accurately patient outcomes is therefore a subject of importance[14] because the physician's expectation of outcome can influence the application and success of treatment.

Method

To study the relationships between prognosis and outcome and between prognosis and cost in primary care, data on prognosis, outcome and cost were collected on all patients who sought care for any acute complaint during a nine-month period at two model family practice clinics at the University of Utah.[15] Less than 1 per cent of the patients refused to participate. Patients who required hospitalization (0.5 per cent), patients who could not be reached for follow-up (8 per cent) and patients who experienced a second, separate episode of illness before follow-up (6 per cent) were excluded from the final sam0le of 1,757 treatment episodes. Return visits for the same problem were included as part of a single episode. This study thus focused on illnesses that had produced a change in function and offered an opportunity to evaluate the consequences of the physician's intervention.

The physicians included first-, second-, and third-year family practice residents and family practice faculty; physician's assistants also saw some patients. All physicians were informed in advance that the study would focus on patient outcome, and their cooperation was solicited. No attempt was made by the investigators to influence physicians' treatment behavior.

Before each patient was seen by the physician, a trained interviewer obtained information about the patient's functional status during the six months prior to the clinic visit and his/her current functional status, reflecting the degree of impairment imposed by the illness. Functional-status ratings were assigned by the interviewer on the basis of the activities reported by the patient. Williamson's seven-level scale was used because it is sensitive to relative function (i.e., it discriminates appropriately for the individual's age and occupation).[16] The categories were: 1) full activity without symptoms; 2) an underlying phys-

TABLE 1. Physician Prediction of Outcome by Actual Outcome

Physician Prediction of Outcome	Actual Outcome		
	Bad	Good	Total
Poor prognosis	56	42	98
Good prognosis	357	1302	1659
Total	413	1344	1757

Sensitivity of prediction $= \frac{56}{413} = 0.136.$

Specificity of prediction $= \frac{1302}{1344} = 0.969.$

Predictive value of a positive test (poor prognosis) $= \frac{1320}{1659} = 0.571.$

Predictive value of a negative test (good prognosis) $= \frac{56}{98} = 0.785.$

ical or laboratory abnormality without symptoms; 3) symptoms, but with full activity; 4) symptoms with restricted activity; 5) limited mobility (i.e., housebound); 6) confinement to bed; and 7) death.

After a patient was seen, the physician was asked to predict 1) the length of time necessary for the patient to receive the maximum benefit from treatment and 2) the functional status the patient should achieve by that time. Any patient whom the physician did not expect to return to his/her usual functional status was defined as having a poor prognosis. The physician was allowed to delay response for up to one week after the patient's first visit in order to make the prognosis on the basis of as much information as possible.

At the time specified by the physician as appropriate for follow-up (not less than two weeks or longer than six months after the visit), the interviewer visited each patient at home to determine his/her actual functional status as a consequence of treatment. The interviewer also obtained any laboratory specimens needed to evaluate the effectiveness of treatment. A good outcome was defined as one in which the patient's functional status on follow-up was equal to or better than his/her usual functional status prior to illness. The number of patients for whom the physicians reported the discovery and persistence of previously unidentified physical or laboratory abnormalities that would prevent a patient from returning to a completely normal functioning status (level 1) was very small (less than 3 per cent).

Physician charges for the office visit and laboratory and x-ray fees were obtained from clinic and hospital billing files. All charges recorded between the initial clinic visit and the follow-up date were included. The physician fee schedule was competitive with that of local private physicians and did not vary according to level of training. Medication costs were determined by the prices charged in the adjoining hospital pharmacy for all medications prescribed by physicians, whether or not the prescriptions were filled. "Other" costs (for services such as physical therapy) were rare; when present, they were included in the total cost and were not examined separately.

Results

The percentage of patients identified by their physicians as having a poor prognosis was very small (6 per cent), as shown in Table 1. In fact, 24 per cent of patients experienced a bad outcome. Although the fraction of patients with a bad outcome was much higher among poor-prognosis patients (56/98), 357 patients with bad outcomes were among the 1,659 patients for whom the physicians expected a good outcome.

These findings can be described in terms of clinical judgment as the predictive value of a positive test (here the proportion of poor-prognosis patients who actually experienced bad outcomes) and the predictive value of a negative test (the proportion of good-prognosis patients who experienced good outcomes), which are here 57 per cent and 79 per cent, respectively. A high predictive value of a positive test is of particular importance in any large-volume operation, such as a mass screening program or primary care itself, in order to ensure that a large number of healthy people are not falsely labeled as ill and subjected to unnecessarily invasive or expensive further testing or treatment. Conversely, a high predictive value of a negative test should protect individuals with true disease from escaping detection and missing appropriate intervention.

Two other parameters frequently mentioned as measures of the usefulness of a screening test or a discriminating variable are the sensitivity and specificity of detecting the disease.[17] In this instance, sensitivity refers to the proportion of bad outcomes that were correctly predicted, and specificity refers to the proportion of good outcomes correctly predicted. The data in Table 1 reveal a sensitivity of 13.6 per cent and a specificity of 96.9 per cent. The crucial difference between the sensitivity and specificity on the one hand, and the predictive value of a positive test and the predictive value of a negative test on the other, is that the first pair refers to the effectiveness or accuracy of a single test in detecting disease and the second pair refers to the consequences of that detection for patients. Both kinds of information about physicians' judgments are important, and the emphasis that may be placed on one or the other relates to the prevalence of the condition under consideration and the costs of applying or missing diagnosis and treatment. In this study we examined the occurrence of a condition (bad outcomes) that was thought to be rare but, in fact, was very common. Therefore, the more frequent error that physicians made was failure to detect patients who would experience bad outcomes.

In order to evaluate the importance of the large number of patients with bad outcomes that were not recognized by their physicians, one can consider how the predictive value of positive and negative tests and the sensitivity and specificity differed by severity of illness, by diagnosis and by the cost of management. Table 2 presents these differences by frequently occurring specific complaints. Problems that a physician might expect to be self-limited (e.g., pharyngitis, upper respiratory infection, flu) had low sensitivities and high specificities, very low predictive value of positive tests and reasonably high predictive value of negative tests. Sensitivity was greater for the more serious problems but was in no case above 40 per cent; however, the predictive value of a positive test increased remarkably for the more serious problems.

Table 3 shows that none of the four measures varied noticeably according to

TABLE 2. Prediction of Outcome for Common Problems

Problem	Sensitivity (Proportion of Bad Outcomes Correctly Predicted)	Specificity (Proportion of Good Outcomes Correctly Predicted)	Predictive Value of a Positive Test (Proportion of Poor Prognoses with Bad Outcomes)	Predictive Value of a Negative Test (Proportion of Good Prognoses with Good Outcomes)	Per Cent of Patients with Bad Outcomes	Number of Patients
Pharyngitis	0.000	0.981	0.000	0.825	17.2	192
Upper respiratory infection	0.000	0.986	0.000	0.802	19.5	261
Otitis media	0.038	0.953	0.167	0.803	19.5	133
Flu	0.000	1.000	—	0.796	20.4	49
Dermatitis	0.000	0.895	0	0.708	26.9	52
Cough and bronchitis	0.074	0.971	0.500	0.725	28.4	95
Neck and back pain	0.375	0.919	0.667	0.773	30.2	53
Abdominal pain	0.258	0.982	0.889	0.701	36.0	86
Headache and malaise	0.267	0.938	0.800	0.577	48.4	62

TABLE 3. Prediction of Outcome by Patient Status at Time of First Visit

Patient Presenting Status	Sensitivity (Proportion of Bad Outcomes Correctly Predicted)	Specificity (Proportion of Good Outcomes Correctly Predicted)	Predictive Value of a Positive Test (Proportion of Poor Prognoses with Bad Outcomes)	Predictive Value of a Negative Test (Proportion of Good Prognoses with Good Outcomes)	Per Cent of Patients with Bad Outcomes	Number of Patients
Asymptomatic but abnormal physical or laboratory examination	0.364	0.941	0.667	0.821	24.4	45
Symptomatic but no limitation of activity	0.134	0.972	0.611	0.773	24.8	993
Symptomatic with minor limitation of activity	0.110	0.969	0.484	0.803	21.1	646
Symptomatic with major limitation of activity	0.200	0.943	0.571	0.758	27.4	73

TABLE 4. Prediction of Outcome by Type of Provider

Type of Provider	Sensitivity (Proportion of Bad Outcomes Correctly Predicted)	Specificity (Proportion of Good Outcomes Correctly Predicted)	Predictive Value of a Positive Test (Proportion of Poor Prognoses with Bad Outcomes)	Predictive Value of a Negative Test (Proportion of Good Prognoses with Good Outcomes)	Per Cent of Patients with Bad Outcomes	Number of Patients
First-year resident	0.090	0.978	0.546	0.785	22.7	295
Second-year resident	0.076	0.963	0.387	0.775	23.3	674
Third-year resident	0.210	0.957	0.618	0.786	24.8	403
Faculty	0.197	0.982	0.824	0.743	29.7	239
Physician's assistant	0.167	0.984	0.600	0.894	12.3	146

the patient's presenting status when ill. Among types of providers (Table 4), third-year residents and faculty demonstrated higher sensitivities and higher predictive value for positive tests than did first- or second-year residents. Physician's assistants predicted bad outcomes more accurately than did first- or second-year residents, but less well than did third-year residents and faculty; this might reflect the influence of supervisory faculty. These differences were small but provide some evidence for a trend toward improvement in accuracy of prognosis with increasing level of training and/or experience.

The long-term economic impact of an unrecognized bad outcome in primary care is not known, but these data can provide some perspective on the cost of a poor prognosis. Differences in cost between the good-prognosis and poor-prognosis patients and between the patients who experienced good and bad outcomes are shown in Tables 5–8. A poor prognosis was associated with higher total costs and increased laboratory and office visit fees. Because a fixed amount ($12) was charged for each office visit, physician office costs indicate the number of patient visits. The dollar differences therefore represent much more laboratory work but only a few more return visits for those patients whose physicians believed that they would not regain their usual functional status. Bad outcomes were also associated with higher laboratory and total costs, but the dollar differences between patients with good and bad outcomes were much less than were those between poor-prognosis and good-prognosis patients. Among patients with a good prognosis, dollar differences in costs between patients with good and bad outcomes were not striking. However, among patients with a poor prognosis, even those who did experience good outcomes had higher costs than did any of the good-prognosis patients, and the increases in total and laboratory costs for poor-prognosis patients with bad outcomes were remarkable. The data thus suggest that prognosis has a much greater effect on the immediate cost of primary care than does outcome.

TABLE 5. Average Costs by Prognosis and by Outcome: Total Costs

	Actual Outcome		
Prognosis	Bad ($)	Good ($)	Overall ($)
Poor	33.05	26.48	30.23
Good	21.36	19.83	20.16
Overall	22.95	20.04	20.72

Discussion

Because a physician's expectation of outcome appears to be an important determinant of cost, it is essential for physicians to become more aware of the factors that influence the cost-generating decisions that are made in the course of a patient's illness. For example, on a patient's first visit, the physician can choose one or a combination of several options: 1) treat the patient on the basis of available information; 2) request additional information, usually in the form of laboratory studies; 3) prolong the period of observation, with or without the initiation of treatment or the gathering of new data, by scheduling a follow-up visit; 4) do nothing more than reassure the patient.

Each of these choices will affect cost and may affect outcome; they must be

TABLE 6. Average Costs by Prognosis and by Outcome: Office Costs

	Actual Outcome		
Prognosis	Bad ($)	Good ($)	Overall ($)
Poor	15.79	13.67	14.88
Good	12.77	12.59	12.63
Overall	13.18	12.62	12.75

TABLE 7. Average Costs by Prognosis and by Outcome: Laboratory Costs

	Actual Outcome		
Prognosis	Bad ($)	Good ($)	Overall ($)
Poor	14.09	7.82	11.42
Good	5.25	3.72	4.05
Overall	6.46	3.85	4.46

evaluated individually in terms of the physician's awareness of risk and benefit to each patient. Unfortunately, our data indicate that primary-care physicians' prognoses may both be incorrect for a large proportion of patients whom they identify as having a poor prognosis (shown by the low predictive value of a positive test) and may also miss a substantial number who fail to return to their usual status after an acute illness (low predictive value of a negative test). Some physicians might comfort themselves that such patients will return voluntarily for further treatment,[18] but we do not know the consequences—physical, psychological or economic—of the intermediate delay.

The fact that physicians did order extra laboratory tests, even during the first visit, for patients with bad outcomes suggests that they may have been uncomfortable about the nature or projected course of certain illnesses in spite of the fact that they predicted good results. However, because a physician in our study could delay making a prognosis until after he/she had received the laboratory reports, the test results must not have been sufficiently indicative of trouble to prompt him/her either to request a follow-up visit or to lower expectation of outcome. Without an interviewer to obtain follow-up data, and with no return visit, the physician might never learn about the patient's remaining impairment. The physician would thus have no regular means of determining whether the investment in the cost of extra laboratory work had produced any benefit. We need much more data before we will know if the additional cost of early identification and treatment of a larger proportion of the primary-care patients who experience bad outcomes offsets the cost of those bad outcomes exerting their effects on function over time, causing patients to seek subsequent care with the same probability of incorrect prognosis.

TABLE 8. Average Costs by Prognosis and by Outcome: Medication Costs

	Actual Outcomes		
Prognosis	Bad ($)	Good ($)	Overall ($)
Poor	3.18	4.62	3.80
Good	3.24	3.17	3.18
Overall	3.23	3.21	3.22

To enable each physician to examine the consequences of his/her actions, we believe that two affirmative steps would be valuable. First, each physician could document the prognosis at each patient visit. Second, each physician could develop an organized system of follow-up, perhaps using inexpensive telephone contacts[19] in order to verify and improve accuracy. This exercise should not only improve the quality of care we give patients, but should also demonstrate our efforts to control costs on an individual basis.

Conclusion

In summary, primary-care physicians' prognostic abilities are imprecise, with low sensitivity and low predictive values for both positive test (poor prognosis) and negative test (good prognosis). During initial contacts with patients, physicians apparently order laboratory tests that they do not pursue with follow-up and do not use to adjust their expectation of patient outcome. The additional investment, therefore, does not seem to assist them in identifying a large number of patients with bad outcomes. A feedback loop is recommended to provide a better perspective on outcome and eventually to improve the efficiency and cost benefit of the medical decision-making process.

References

1. Draper P, Smits HL. The primary-care practitioner: Specialist or jack-of-all-trades. N Engl J Med 1975;293:903.

2. Garfield SR, Collen MF, Feldman R, et al. Evaluation of an ambulatory medical-care delivery system. N Engl J Med 1976;294:426.

3. Eisenberg JM, Rosoff AJ. Physician responsibility for the cost of unnecessary medical services. N Engl J Med 1978;299:76.

4. Wright DD, Kane RL, Snell GF, et al. Costs and outcomes for different primary care providers. JAMA 1977;238:46.

5. Pauker SG, Kassirer JP. Therapeutic decision making: A cost benefit analysis. N Engl J Med 1975;293:229.

6. Schwartz WB, Gorry GA, Kassirer JP, et al. Decision analysis and clinical judgment. Am J Med 1973;55:459.

7. Shapiro AR. The evaluation of clinical predictions. N Engl J Med 1977;296:1509.

8. Peterson WL, Sturdevant RAL, Frank HS, et al. Healing of duodenal ulcer with antacid regimen. N Engl J Med 1977;297:341.

9. Gorry GA, Pauker SG, Schwartz WB. The diagnostic importance of the normal finding. N Engl J Med 1978;298:486.

10. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926.

11. Daniels M, Schroeder SA. Variation among physicians in use of laboratory tests. II. Relation to clinical productivity and outcomes of care. Med Care 1977;15:482.

12. Sturdevant RAL, Stern D. Accuracy of physicians' prediction of cholecystography results. Med Care 1977,15:488.

13. Mushlin AI, Appel F, Pozen JT, et al. The use of patient-reported problem status in evaluating ambulatory care. Clin Res 1975;23:267.

14. Koran LM. The reliability of clinical methods, data and judgments, Parts I and II. N Engl J Med 1975;293:642.

15. Kane RL, Gardner J, Wright DD, et al. A method for assessing the outcome of acute primary care. J Fam Pract 1977;4:1119.

16. Williamson JW. Outcomes of health care: Key to health improvement. In: Hopkins CE. Methodology of identifying, measuring and evaluating outcomes of health service programs, systems and subsystems. Washington, D.C.: Government Printing Office, 1977.

17. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. N Engl J Med 1975;293:211.

18. Mushlin AI, Appel FA. Diagnosing potential noncompliance: Physicians' ability in a behavioral dimension of medical care. Arch Intern Med 1977;137:318.

19. Henk M, Froom J. Outreach by primary-care physicians. JAMA 1975;233:256.

MEDICAL CARE
February 1982, Vol. XX, No. 2

Hospital Discharge Queues in Massachusetts

LEONARD W. GRUENBERG, PH.D., AND THOMAS R. WILLEMAIN, PH.D.

There is growing concern over the inappropriate utilization of health care facilities. The high cost of hospital care and the apparent shortage of nursing home beds have focused attention on one aspect of this problem: the clinically unnecessary days (sometimes called "administrative days" or "ADs") spent in hospitals by patients who are awaiting placement in long-term care facilities. In this study, data from a 1976 Massachusetts Department of Public Health survey of patients backed up in hospitals were analyzed to determine the magnitude of the problem and to examine the influence of several major factors that had been hypothesized in previous studies to contribute to the backup. We demonstrate that the average delay of a patient found waiting in a "snapshot" survey (which is often used to estimate the magnitude of the problem) is significantly greater than the average delay experienced by a typical discharged patient. We show that there are at least two major factors that influence the delay time: nursing home preferences in accepting certain types of patients and nursing home occupancy rates in the hospital service area. Neither medical-surgical occupancy rate nor the number of AD patients waiting in the hospital was significantly correlated with the delay time.

SPIRALING HEALTH CARE COSTS have sparked growing concern over the inappropriate utilization of health care facilities. The high cost of hospital care, as well as the apparent simplicity of a number of proposed remedies, has focused attention on one aspect of this problem—an aspect that is known as the problem of "administrative days."

Administrative days are a phenomenon of the juncture between acute and long-term care. Acute hospital patients with chronic disease or disability who require less intense but more prolonged treatment sometimes find their hospital discharges delayed because a suitable placement has not been arranged. The days spent in hospitals merely awaiting placement are referred to as "administrative days" (ADs) to distinguish them from that portion of the length of stay deemed clinically appropriate.

The purpose of this article is to examine the extent to which certain factors appear to influence the length of time it takes for individuals in hospitals to be placed in long-term care facilities. The factors that we have studied as possible determinants of the delay—patient characteristics, hospital occupancy rate and area wide nursing home occupancy rates—were chosen because they relate directly to solutions that have been proposed to minimize the undesirable effects of the AD problem.

From the University Health Policy Consortium, Brandeis University, Waltham, Massachusetts.

Supported in part by a Health Care Financing Administration grant to the University Health Policy Consortium, #18-P-97038/1-03.

Reprint requests: Leonard Gruenberg, UHPC, Heller School, Brandeis University, Waltham, MA 02254.

0025-7079/82/0200/0188/$01.20 © J. B. Lippincott Co.

Administrative days may be perceived either as a problem in themselves or as a symptom of other problems. Those concerned about the level of hospital expenditures object to ADs on grounds of cost. Those concerned about overmedicalization of long-term care object to the extension of hospital stay on grounds of quality of care and quality of life. Others see in ADs a reflection of some other health planning issue: nursing home bed shortages (which might create backlogs), low hospital occupancy rates (which might tempt hospitals to prolong stays), or discrimination against certain classes of patients (which might be interpreted as either venal or prudent, depending on further analysis of causes).

Whether regarded primarily as a problem in themselves or as evidence of more basic difficulties, ADs have been attracting the attention of both the acute and long-term care communities for some time, and interest is growing. In Massachusetts the state chapter of the Society for Hospital Social Work Directors of the American Hospital Association conducted two studies (1972 and 1973); the Massachusetts Hospital Association conducted three studies (1974, 1976 and 1980); the state Department of Public Health conducted three (1974, 1976 and 1980); and the state Department of Public Welfare produced two studies in 1980 alone. In New York, the state Council of Professional Standards Review Organizations undertook a study of the AD problem in 1979 and followed with another in 1980, labeling the problem a "crisis." Shapiro, Ross and Kavanaugh[1] documented the extent of the problem in Manitoba, Canada. The Office of the Inspector General, Department of Health and Human Services[2] studied the AD problem in several states at the request of the Secretary and predicted that it would grow more serious.

Scanning these studies, one discovers two sets of issues. The first concerns the estimation of the *magnitude* of the AD problem; the second concerns its *explanation*. Every AD study we have read has been based primarily on a one-day "snapshot" survey designed to estimate the magnitude of the problem at a particular time. Such snapshot surveys introduce distortions that appear not to be widely appreciated, so we will discuss the issue of distortion. Our main task, however, is to help with the issue of explanation by testing three different interpretations of the AD phenomenon against data from a study conducted in 1976 by the Massachusetts Department of Public Health.

The three competing explanations have been distilled from the excellent summary paper by Beatrice.[3] One view holds that ADs arise because nursing homes discriminate against certain classes of patients, being reluctant to admit those whose care will be paid for by Medicaid and especially reluctant to admit Medicaid patients with "heavy care" needs and/or behavior problems. A second view holds that ADs arise because there are too few nursing home beds and/or home care services to accommodate the growing number of frail elderly, thereby forcing backlogs in the hospitals. The third view holds that hospitals are responsible for ADs, either because AD patients represent "easy money" when the hospital is faced with a low occupancy rate or because the hospital discharge process is inefficient.

All three theories have received some support in the literature. The 1979 Massachusetts Hospital Association study[1] concluded that the supply of nursing home beds was too low in general and that nursing homes were reluctant to admit Medicaid and/or "heavy care" patients. As might be expected, the Hospital Association report stressed the diligence of hospitals in seeking placement. On the other hand, the 1980 Massachusetts Department of Public Welfare study[5] declared that patient characteristics played a very minor role in the AD problem as compared with disinterested or incompetent placement efforts by hospitals. The Department's study agreed that nursing home bed shortages exacerbated the problem but also noted patterns of "sub-area politics" that restricted the effective bed supply to a number below that nominally available. Nationally, the Inspector General's report[2] cited evidence for all three theories: nursing home selectivity, hospital incentives to keep AD patients and nursing home bed shortages. In this study we first take a fresh look at the magnitude of the AD problem, then examine the evidence for the three competing explanations. We also draw inferences about useful responses.

Data

The data set used to study the magnitude and determinants of the AD problem is a survey conducted by the Massachusetts Department of Public Health (MDPH) on March 10, 1976 and followed up on April 19, 1976.[6] The survey was a repeat of an earlier one carried out in 1974.[7] A copy of the survey instrument is available from the authors. Respondents were discharge planners in each of the acute and chronic disease/rehabilitation hospitals in Massachusetts. Each was asked to report on those patients who were awaiting discharge to a lower level of care on the morning of the survey. The follow-up survey six weeks later was to ascertain the disposition of each case. A total of 1,322 individuals were found to be awaiting placement on March 10, 1976. Of these, we will focus only on the 620* adults who were waiting in acute hospitals for placement in a skilled nursing facility (SNF) or an intermediate care facility (ICF), since it is the acute hospital/nursing home interface that is the subject of this study.

The survey provided two different measures of ADs for each patient. One was the time the patient had spent on AD status up to the day of the survey. The other was the time from the survey date until the patient left the AD queue, either because she was placed, died, or again required the clinical services of the hospital; for the 12% of our patients whose delay exceeded the six-week follow-up period, the exact duration could not be recorded. (We conservatively recorded these stays as 43 days.) We took as the estimate of ADs for each patient the sum of the periods before and after the survey date.

The survey recorded roughly 30 patient descriptors, including demographics, measures of dependence in activities of daily living (ADL), rehabilitation potential, behavior problems, types of service required, level of nursing home care needed and sources of payment for care. The survey also recorded the identification number and health systems agency of the hospital in which the patient waited. These identifiers allowed linkage to other data bases on hospital and nursing home occupancy.

Estimation of ADs

The procedure usually used to estimate the magnitude of the AD problem has an obvious advantage and a less well-known disadvantage. That procedure is to obtain a "snapshot" of the AD problem on a particular day. A snapshot survey is economical and practical, since effort—and compliance—need only be mustered for a single day's data collection. To the extent that the chosen day is typical, the snapshot provides a good estimate of the number of patients awaiting placement.

The disadvantage of a snapshot survey is that if the data are used to estimate the delay of a "typical" patient who becomes ready for discharge on a given day, the resulting estimate is severely biased: the average delay of persons found in the snapshot is significantly greater than the average delay of a typical person being discharged. The bias arises because patients with longer delays are more likely to be found waiting on any particular day than are patients with brief delays. Many of the latter can come and go in the system without detection by a snapshot. In contrast, patients with long delays are at proportionately greater risk of discovery. Of course, an additional problem with snapshots is that only that part of the total delay that has already transpired can be recorded; a follow-up survey is needed to determine the total delay for each patient.

---

* Of these 620 patients, 45 lacked a full set of data on personal characteristics. These missing cases had very long waiting times (mean = 156 days, SD = 347 days). Most of the analyses focus on the 575 patients for whom a full description was available.

When certain mathematical conditions are satisfied, the theory of "length-biased sampling"[8] or "random incidence"[9] can be used to convert the data on partial delays found in a snapshot into unbiased estimates of total delay as found in a longitudinal study. One key result of this theory is that the mean interval measured in a snapshot survey is a multiple of the true mean number of ADs. The multiplier involves the degree of variability in the distribution of ADs: the more variability, the greater the overestimate obtained from a snapshot. It can even happen that one group of patients with a shorter mean placement delay than a second group can have a longer mean delay in a snapshot; this would arise if the first group contained a significant minority of individuals with very long delays and the second group did not.

That the 1976 MDPH survey included a follow-up after six weeks allows a demonstration of the magnitude of the overestimate of placement delay produced by a snapshot. Since even the six-week follow-up period was not long enough to capture the full delay of many patients, estimates of mean delay will be distorted; we will therefore cite median delays instead. Consider adult, acute hospital patients awaiting placement in a nursing home. Of 620 such patients, 102 became ready for discharge on the day of the snapshot survey ("new patients"), and 518 were found already waiting ("old patients"). The median total wait for the old patients was 21 days, whereas the median for new patients was only five days. This fourfold difference should be significant for discussions of the magnitude and nature of the AD problem. As we note below, the same phenomenon appeared in the utilization of nursing homes; according to results obtained from the 1977 National Nursing Home Survey,[10] the median length of stay was less than three months among nursing home discharges but greater than 12 months among nursing home residents.

Having noted these distortions in estimates of delay, we are nonetheless forced to use the survey figures in our subsequent analyses. Since these analyses concern comparative rather than absolute magnitudes, the data suffice for our purposes. The reader is cautioned to remember, though, that the mean delays we cite are inflated by the length-biased sampling of the survey.

Explanation of ADs

We gained considerable insight into the waiting time problem by examining the distribution of waiting times. Table 1 shows how the total number of AD patients from the entire group ("new" and "old" patients) who were found to be still awaiting placement decreased with time after the date of the survey. It can be seen that the number waiting drops sharply during the first few days after the survey: in fact, of the 620 patients who were found to be waiting on the day of the survey, 156 (slightly more than one quarter) were placed within two days. On the other hand, the distribution has a long "tail." Observe in Table 1, for example, that one quarter of the patients were still waiting after three

TABLE 1. Per Cent of Patients Remaining in Discharge Queue after Survey Date

Weeks after Survey	Per Cent Remaining
1	51.9
2	34.7
3	25.5
4	17.9
5	13.9
6	11.6

weeks and 12 per cent were still not placed when the follow-up was made six weeks after the survey date.

This analysis suggests that there are (at least) two types of patients: those "easy to place" and those "hard to place." However, the considerable spread in the distribution points to the fact that these two groups may not be very well defined, i.e., that some patients who look quite similar to the patients in the hard-to-place group may have been placed in a short period of time and vice versa.

To proceed with the analysis, half of the data (every second case) was first set aside to serve as a confirmatory data set. The remaining cases served as an exploratory data set on which to develop hypotheses [illegible] models.

Patient Characteristics

The first step in the analysis was to look at two groups of patients who represented extremely short and extremely long stays. Included were 28 patients whose total waiting time (i.e., the time interval between when they became ready and when they were placed) was two days or less and 115 whose wait exceeded 30 days. We found major differences between the two groups on several variables. Patients in the "easy" or short-stay group were found to be more likely to need physical therapy, to be rehabilitatable and likely to return to the community, to have a payment source other than Medicaid, and to be admitted to the hospital from home or from a nursing home (as opposed to from some other institutional or group setting). Patients in the long-stay group were more likely to be totally dependent if they were to return home; their household was most often regarded as unable to cope physically and emotionally with the patient, and they were more likely to have behavior problems. This analysis of cases with extremely long or short stays, together with other multivariate analysis and anecdotal descriptions of the system provided by Welfare Department officials and discharge planners, led us to examine groupings or profiles of patients using the following variables as the building blocks of the profiles:

1. Source of Admission to the Hospital—Patients were grouped into three subgroupings: admitted to the hospital from home, from a nursing home, or from some other institutional or group setting (labeled as "other").

2. Likely Source of Payment in the Nursing Home—Patients were classified as either Medicaid or non-Medicaid (the latter being mostly Medicare or self-pay).

3. Likelihood of Returning to the Community—Patients were classified into two subgroups: those who were likely to return, and all others (this included the group who were in need of terminal care, as well as those who might return and those who were thought definitely to require permanent institutionalization).

4. ADL—The patients were classified into two groups: good ADL and poor ADL. The poor ADL group were dependent on at least five of the following six functions: bathing, dressing, toileting, transferring, continence and feeding. According to the work of Katz,[11] the most likely distinction between the two groups was continence, i.e., patients in the good ADL group were likely to be continent while patients in the poor ADL group were likely to be incontinent. We thus label the poor ADL group "incontinent."

5. Behavior—Patients were classified into two groups: those exhibiting frequent behavior problems of any type and those

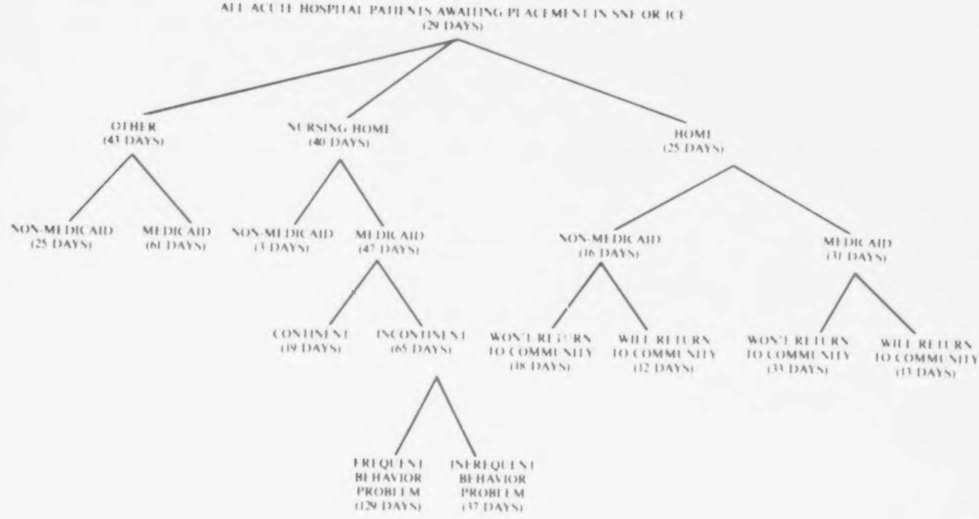

ALL ACUTE HOSPITAL PATIENTS AWAITING PLACEMENT IN SNF OR ICF
(29 DAYS)
OTHER
(43 DAYS)
NURSING HOME
(40 DAYS)
HOME
(25 DAYS)
NON-MEDICAID
(25 DAYS)
MEDICAID
(61 DAYS)
NON-MEDICAID
(3 DAYS)
MEDICAID
(47 DAYS)
NON-MEDICAID
(16 DAYS)
MEDICAID
(31 DAYS)
CONTINENT
(19 DAYS)
INCONTINENT
(65 DAYS)
WON'T RETURN
TO COMMUNITY
(18 DAYS)
WILL RETURN
TO COMMUNITY
(12 DAYS)
WON'T RETURN
TO COMMUNITY
(33 DAYS)
WILL RETURN
TO COMMUNITY
(13 DAYS)
FREQUENT
BEHAVIOR
PROBLEM
(129 DAYS)
INFREQUENT
BEHAVIOR
PROBLEM
(37 DAYS)

FIG. 1.

exhibiting no or infrequent behavior problems.

Using these five variables, we created ten distinct subgroups. The findings regarding the mean waiting time of these different subgroups in the exploratory sample are presented in Figure 1. The major findings are as follows:

1. For each referral source, there were striking differences between the mean waiting times of patients who would have different sources of payment in the nursing homes. The differences are most pronounced for the patients admitted to the hospital from nursing homes—Medicaid patients had to wait 47 days before being placed, while non-Medicaid patients were placed after an average of three days. Patients admitted to the hospital from home or from an "other" referral source also had a much longer delay in the hospital (roughly twice as long) if their stay in the nursing home was to be reimbursed by Medicaid.

2. For Medicaid patients admitted from nursing homes, there were two significant distinctions: patients who were incontinent took much longer to place than those who were continent; and among those who were incontinent, those with frequent behavior problems were much more difficult to place than those who did not have such problems.

3. For patients admitted from home, there was a common pattern for Medicaid and non-Medicaid patients: individuals who were likely to return to the community were placed much more rapidly than others.

After completing the analysis of the exploratory data set, we tested the reproducibility of these findings by examining the confirmatory data set. First, we studied the relatively refined distinctions between pairs of subgroups as shown in the lower "branches" of Figure 1 by determining whether the confirmatory data set yielded the same relative order for the mean waiting times of those pairs of subgroups in the lower branches. We found that in the case of one of these "contrasts," that between the two behavioral subgroups who were part of the nursing home–Medicaid–incontinent group, the mean waiting times reversed order in the confirmatory data set. For this reason, we collapsed these two subgroups into one.

After completing this analysis, we combined the exploratory and confirmatory data sets to test the significance of the observed differences in means. When differences were not significant, we merged the subgroups. In this way, the number of groups was reduced to five. The mean value and standard deviation of the waiting time of each of the five distinct subgroups are shown in Table 2.

Discussion of Analysis of Patient Characteristics

How Much Can Patient Characteristics Explain? Table 2 shows that it is possible to identify subgroups of patients with different profiles who will experience very different delay patterns. Moreover, the distinctions among these subgroups hold up when the data are subjected to a confirmatory analysis. This strongly suggests that the subgroups defined in Table 2 have a real meaning in that they relate to the preferences of nursing homes for different types of patients.

At the same time, there remains quite a bit of unexplained variation in the waiting times, as evidenced by the rather large standard deviations. Note that the standard deviation of the total waiting time of each subgroup in Table 2 is very close in magnitude to the mean value. This feature is characteristic of the geometric probability distribution.†

This suggests that although there are distinct subgroups of patients having very different expected waiting times, there is a great deal of variation in the amount of time any particular type of patient must wait. This random variation is not related to the patient characteristics we studied but is perhaps connected with other variables (hospital, area, nursing home) not included in this part of the analysis.

It is also possible to make quantitative estimates of the explanatory power of the patient characteristic variables. This was done by conducting an analysis of variance of differences in the total waiting time that can be explained by group membership. The ten-group model accounted for 27 per cent of the variance in the exploratory sample and for 11 per cent in the confirmatory sample. This dramatic drop attests to the wisdom of employing a separate confirmatory sample. The five-group model accounted for 12 per cent of the variance in the combined data sets. By conventional standards, accounting for 12 per cent of the variance is not a major accomplishment. However, the special nature of the placement process makes the 12 per cent figure respectable. In ordinary regression analysis or analysis of variance, the underlying theoretical model assumes that the error has a Gaussian distribution with constant variance throughout the range of the dependent variable. However, as we have already noted, waiting time is well described as a geometric random variable, for which the standard deviation nearly equals the mean. Thus, the percentage of variation accounted for depends strongly on the predicted mean waiting time, and there is a large, irreducible component of random error even for a model that classifies patients and estimates means perfectly. For instance, suppose that the five groups from Table 2 were in fact the basic subdivisions of the AD population and their relative proportions and mean waits were exactly as estimated. We show in Appendix B that, in this case, a perfect model would account for only about 15 per cent of the variance in waiting times. Thus, the 12 per cent figure represents about 80 per cent of the maximum achievable explanatory power.

Who Gets Placed and Who Must Wait? Additional insight into this question can be

---

† The geometric distribution describes the situation that occurs when dice are rolled. For a single die, the probability p of any particular number coming up is 1/6. The probability that you will throw the die n times without getting that number is given by (1-1/6) to the n^{th} power.

TABLE 2. Mean and Standard Deviation of the Total Waiting Time for the Five Distinct Patient Subgroups

Description of Subgroup	Mean	Standard Deviation	Number of Patients
A. Nursing home, non-Medicaid	3.3	3.1	15
B. Home, will return to the community, Medicaid and non-Medicaid	11.8	13.1	74
C. Other, non-Medicaid; other, Medicaid; nursing home, Medicaid, continent; home, non-Medicaid, unlikely to return to community	18.7	19.1	189
D. Home, Medicaid, unlikely to return to the community	38.4	45.9	213
E. Nursing Home, Medicaid, incontinent	50.1	58.7	84
Total	29.3	38.0	575

gained by further collapsing the five groups of patients into two larger groups, which we will call "easy" (for easy to place) and "hard" (for hard to place). This simplification costs little in predictive power, since the two-group model accounts for 10 per cent of the variance in the combined data sets, compared with 12 per cent for the five-group model. The 10 per cent figure represents about 70 per cent of the systematic variation that can be explained. The "hard" group comprises subgroups D and E in Table 2; these Medicaid beneficiaries have mean waiting times greater than the overall mean. Most of these "hard" patients (213 of 297) are individuals who were residing in the community prior to their hospitalization. They are regarded by the discharge planners as unlikely to return to the community, i.e., they are likely to become permanent nursing home patients. The remaining 84 patients who are "hard" are often called the "heavy care" nursing home patients. They are patients who were sent to the hospital from a nursing home; they are very dependent in ADL (incontinent) and paid for in the nursing home by Medicaid. After their hospitalization, the nursing homes do not want to take them back.

All other patients (identified as subgroups A, B and C in Table 2) are in the "easy" group, which includes all of the non-Medicaid and the remaining Medicaid patients. Note, though, that it is quite possible for an "easy" patient to endure a fairly long wait. The easy/hard distinction refers to average waiting time.

As we observed earlier (see section on estimation of ADs), the patients who are found in the queue present a biased picture of the waiting time problem that does not represent the situation of a typical hospital patient who becomes ready for discharge to a nursing home. It is very instructive to examine the composition (in terms of "easies" and "hards") of these two cross sections of patients. The numbers of patients and mean waiting times of these two cross sections are shown in Table 3. In this table, the "new" patients are those who became ready on the day of the survey, and the "old" patients are those who became ready at some earlier date and were already in the queue on the day of the survey.

The following observations are based on the figures presented in Table 3:

1. "New" patients are much more likely to be "easy" than "hard." The "easies"

TABLE 3. Mean Waiting Times of Easy and Hard Patients for New and Old Cross Sections

Patient Type	Cross Section New	Cross Section Old	All Patients
Easy			
No. and % of patients	69 (71%)	209 (44%)	278 (48%)
Mean waiting time (days)	6.2	19.2	16.0
Hard			
No. and % of patients	28 (29%)	269 (56%)	297 (52%)
Mean waiting time (days)	13.4	44.8	41.8
Total			
No. and % of patients	97 (100%)	478 (100%)	575 (100%)
Mean waiting time (days)	8.2	33.6	29.3

make up 71 per cent of the "new" group but only 44 per cent of the "old" group.

2. The waiting times of both "easy" and "hard" patients are more than three times greater in the "old" group than in the "new" group.

These observations provide a graphic demonstration of our claim that a snapshot survey exaggerates the length of time spent on AD status. People already in the queue are much more likely to be "hard" than "easy" and to have longer waiting times than the "new" patients who become ready on a given day.

It is also instructive to examine the differences between "easy" and "hard" to place patients against the backdrop of a snapshot photograph of patients who already reside in nursing homes. As we noted in the section on estimation of ADs, there is a significant difference in expected length of stay between a typical patient admitted to a nursing home and a typical resident of a nursing home. A comparison of the distribution of length of stay of these two groups is shown in Table 4. These data are from the 1977 National Nursing Home Survey.[10]

Table 4 shows that nearly two thirds of the discharged patients have a stay of less than six months; on the other hand, nearly three quarters of resident patients have been in the nursing home for more than six months. It is unlikely that any patient who has resided in a nursing home for more than six months is going to be discharged home.‡ Thus we can see that the bulk of *residents* of nursing homes are of the long-stay, custodial type. Most of these patients are paid for by the Medicaid program. Thus, it appears that the most typical nursing home *resident* (as distinct from admission)—the long-stay custodial care patient—is the type of patient nursing homes are least likely to want and is the type most likely to be backed up in the hospital.

Hospital Characteristics

We investigated the role of hospitals themselves in influencing ADs by correlating hospital characteristics with residual waiting times, i.e., with the differences between observed waiting times and waits expected under the two-group model. We used residuals rather than actual waiting

---

‡ In fact, according to data obtained from the National Nursing Home Survey, whereas 36 per cent of patients who are discharged from nursing homes after a stay of six months or less are discharged to their own or their family's home, less that 4 per cent of those patients discharged after more than six months return home.

TABLE 4. Comparison of Lengths of Stay of Residents and Discharged Nursing Home Patients, U.S., 1977

Length of Stay	Resident Patients	Discharged Patients
Less than 3 months	14.5%	53.6%
3–6 months	9.4%	12.0%
6–12 months	12.5%	10.0%
1–3 years	32.8%	14.0%
3–5 years	14.8%	6.1%
5 years or more	16.0%	4.1%
Total	100.0%	100.0%

times to provide some statistical adjustment for the varying mix of "easy" and "hard" patients across hospitals. The hospital characteristics we used were the size-adjusted occupancy rate and the number of patients found waiting by the survey. We chose these characteristics to test two theories about hospitals: that the AD problem will be worse in hospitals with low occupancy rates and that hospitals with many AD patients are "bad actors."

In the analysis using occupancy rate, we divided the 93 hospitals in the exploratory sample into six bed-size groupings. We then classified the hospitals into two groups, according to whether their occupancy rates were above or below the average for their bed-size category. This standardization was necessary because the occupancy rate has little normative significance unless related to bed size. Mean residual waits were three days longer in hospitals with below-average occupancy, but there was so much overlap between the two groups that the difference in mean residuals had no statistical significance ($Z = 0.8$, $p > 0.40$). If hospitals with low occupancy relative to their peers feel pressures to "hold on" to AD patients, the effect is not large enough to be reliably detected in our data set.

The second analysis focused on the relationship between mean residual waiting times and number of AD patients found in a given hospital on the survey day. The analysis revealed that the presence of large numbers of AD patients is not itself a sign of hospital complicity in the AD problem. A hospital with a large number of AD patients was not more likely to have a higher-than-average wait, after adjusting for the case mix, than a hospital with only a few AD patients (rank order correlation = 0.05, $P > 0.49$).

These findings do not necessarily mean that hospitals share none of the blame for the AD problem. The Massachusetts Department of Public Welfare[5] documented cases of halfhearted or inefficient placement efforts in a study of 14 Massachusetts hospitals. Nevertheless, it appears that neither low occupancy nor large numbers of AD patients are reliable indicators of hospital complicity.

HSA Characteristics

Besides patient and hospital characteristics, characteristics of the HSA in which the hospital is located might influence waiting times. Residual waiting times did vary by HSA, as shown in Table 5. Since the highest and lowest mean residual waits occurred in the HSAs with the highest and lowest combined occupancy rates for skilled and intermediate level facilities in 1975, it was natural to explore the relationship between HSA nursing home occupancy and residual waiting time. Our analysis used weighted regression methods to take account of differences among HSAs in the reliability of estimates of mean residual wait. We found a positive relationship between occupancy and residual waiting time:

$$\text{Residual Wait} = -462.3 + 4.922 \text{ Occupancy Rate}, \quad R^2 = 0.65,\ F = 7.58,$$

But since this positive relationship depends so heavily on the two extreme

TABLE 5. Residual Wait and Nursing Home Occupancy by Health Service Area

	Residual Wait			
HSA	Mean	SD	Number of Patients	Nursing Home Occupancy
1	1.9452	37.8119	79	92.28
2	−4.8763	27.4472	81	93.66
3	2.5614	22.6969	41	93.54
4	−1.9277	38.3576	198	94.25
5	−13.0653	17.8162	61	91.65
6	11.4344	51.0310	115	94.68
Total	0.000	37.3902	575	

points, it should be viewed with some caution.

To get a more robust estimate of the relationship, we also computed a "jackknife" estimate,[12] which is based on calculations requiring the sequential deletion of each HSA from the data set. The jackknife estimate is calculated as follows:

$$\text{Residual Wait} = -678.5 + 7.215 \text{ Occupancy rate} \atop (1.88)$$

where the approximate t-statistic is shown in parentheses. This analysis, which is a more realistic depiction of a replication of our analysis, shows a positive relationship of marginal statistical significance.

It might be that some other characteristic of the HSAs is more fundamentally related to waiting time, with occupancy only a proxy. However, nursing home occupancy has an obvious a priori appeal and did relate positively to residual waiting time. Since nursing home occupancy across the HSA in which the hospital is located is a blunt indicator of nursing home bed availability, analyses using it will tend to underestimate the correlation with residual waiting time. Perhaps subsequent analyses will take more careful note of localized occupancy variations within HSAs and thereby discover an even stronger relationship.

Conclusions

This analysis suggests that the roots of the AD problem are not simple. In fact, our findings support two hypotheses: 1) as indicated by the sharp differences in the average waiting times of the different patient types, nursing homes are indeed admitting patients very selectively; and 2) the dependence of residual mean waiting time (i.e., mean waiting time after controlling for differences in patient-mix) on geographic area is an indication that the supply of beds has an effect on waiting times.

In addition, it is likely that the third explanation—that certain hospitals are dragging their feet—may also be correct. Although we have found that neither hospital occupancy rates nor the absolute number of patients waiting in the hospital correlates with the case-mix adjusted mean waiting time, the existence of large residual mean waiting times for a number of hospitals strongly suggests that sluggish hospital discharge planning is probably a factor in certain hospitals. What our results fail to establish is an association between foot-dragging and either low occupancy or large numbers of AD patients.

The results of this analysis, coupled with some experience with organization of long-term institutional services, can help in developing strategies for reducing the undesirable effects of the AD problem.

We first note that our analysis leads to the rejection of some of the simpler proposals. For example, one such proposal is to expand the supply of beds. Although a significant expansion in bed supply might help to alleviate the problem, this would not be an efficient solution; the sharply different mean waiting times for different types of patients suggest that it would take a very large increase in bed supply to reduce the AD problem significantly. Whatever improvement in access would occur is likely to benefit those whose access is already best.

Similarly, proposals to penalize hospitals by significantly reducing their reimbursement for ADs will have limited impact, given the extreme reluctance of nursing homes to accept certain types of patients.

The strategy of paying nursing homes more to take care of certain types of patients (the so-called "heavy care" patient) may lead to a reduction in ADs, provided the change in reimbursement is sufficient to overcome providers' reluctance to accept certain types of patients. But note that differential reimbursement probably requires a simultaneous change in regulatory requirements regarding nursing home levels of care. (This is discussed in more detail in the "Discussion" section.) At the very least, a reimbursement-based initiative must take into account the fact that many of the "undesirables" who are backed up in the hospitals have similar profiles to many (if not the majority of) patients who are already residing in nursing homes.§

Discussion

The current regulatory system, which attempts to encourage nursing homes to group patients according to the level of nursing skill required in their treatment, ignores the very important differences among individuals in the type and total amount of care they need.[13] This has several important consequences, which exacerbate the AD problem. First, it must be noted that each level of care includes patients with a broad range of disabilities requiring widely disparate amounts of care. Nursing homes tend to select those patients who require a smaller total amount of care, all other things being equal. For this behavior, often called "creaming," providers are accused of being self-seeking and profit oriented. However, this behavior could be interpreted as a rational response to the regulatory/reimbursement system. Since most custodial long-term patients will require more care as time progresses, and since nursing homes are forced by reimbursement practices to work with fixed staffing patterns, the homes select less disabled new patients to balance an increasingly sick case load.

A second consequence of the "level of skill" definition is that it has fostered the mixing, in single nursing units, of patients having diverse and difficult problems—including dementia, frequent incontinence and various behavioral disorders—with patients whose primary limitations are physical. It is not surprising that many nursing homes are reluctant to accept patients who have special problems; it is, in fact, very difficult for both the staff and for the physically handicapped but mentally alert patient to deal with the problems caused by these difficult patients.

---

§ The state of Massachusetts is currently addressing the AD problem by providing augmented Medicaid reimbursement rates for nursing homes who accept AD patients. This policy does not take into account the similarity in profile of the typical AD patient and the typical nursing home resident. This policy appears to us to be ill-advised; it will create an incentive for nursing homes to refer more of their patients to acute hospitals, since, when these patients return, the nursing home may in some circumstances be paid at a higher rate. Furthermore, since the policy does not distinguish among patients by length of time on AD status, it encourages creation of ADs for easy-to-place patients, while offering no differential reward for nursing homes to admit hard-to-place patients.

We believe, therefore, that current regulations that differentiate among patients according to the level of nursing skill needed in their care are not compatible with a rational organization of services in long-term care facilities. Modifications in level-of-care regulations and in reimbursement practices may facilitate a solution to the AD problem. The new regulations should allow facilities to group patients according to the total amount of care they need (instead of the level of skill needed) and to segregate patients having certain types of conditions into separate units (with separate reimbursement). A system such as this has been developed and successfully implemented in at least one large long-term care facility,[14,15] and the time appears right for a larger-scale experimentation with these concepts.

References

1. Shapiro E, Roos NP, Kavanaugh S. Long term patients in acute care beds: Is there a cure? The Gerontologist 1980;20:342.

2. U.S. Department of Health and Human Services, Office of the Inspector General. Secretarial report, Restricted patient admittance to nursing homes: an assessment of hospital back-up. August 1980.

3. Beatrice D. Briefing paper: Administratively necessary days. Internal report prepared for the Massachusetts Department of Public Welfare, December 1979.

4. Dumbaugh KA, Mackler RO. Report on patients in Massachusetts hospitals awaiting placement into long-term care. Massachusetts Hospital Association, October 1979.

5. Massachusetts Department of Public Welfare, Long Term Care Unit. The administrative day problem: an analysis of selected hospitals. July 1980.

6. Jordan H, Gruenberg LW. Trends in the problem of placing Massachusetts hospital patients in long-term care facilities. Presented at the 104th Annual Meeting of the American Public Health Association, October 1976.

7. Beattie RT, Jordan HS. A preliminary analysis of a survey of Massachusetts hospital patients who were ready for discharge and awaiting placement in other facilites on March 20, 1974. Massachusetts Department of Public Health Working Paper, April 1975.

8. Cox DR, Lewis PAW. The Statistical Analysis of Series of Events. New York: Wiley, 1966, 61.

9. Drake AW. Fundamentals of Applied Probability Theory. New York: McGraw-Hill, 1967;149.

10. National Center for Health Statistics. The national nursing home survey: 1977 summary for the United States. July 1979.

11. Katz S, et al. Studies of illness in the aged, the index of ADL: A standardized measure of biological and psychosocial function. JAMA 1963.

12. Mosteller F, Tukey JW. Data Analysis and Regression: A Second Course in Statistics. Reading, Mass.: Addison-Wesley, 1977;133.

13. Bishop CB, Plough AL, Willemain TR. Nursing home levels of care: Problems and alternatives. Health Care Financing Review 1980;2:33.

14. Besdine RW, Gruenberg LW, Kierman M, May MI. A cost-effective quality of life level of care system for the institutionalized elderly. Presented at the 107th Annual Meeting of the American Public Health Association, October 1980.

15. Besdine RW, Gruenberg LW, Frady L, May MI. A stratified care system promoting high quality, cost-effective institutional long-term care. Presented at the 108th Annual Meeting of the American Public Health Association, October 1980.

Appendix: Analysis of Mixtures of Geometric Distributions

We consider the magnitudes of the coefficient of determination (R^2) that will arise in regression analysis of waiting time distributions formed by mixtures of geometric random variables. Suppose that there are G groups, each containing N_i patients and having a geometric distribution of waiting time with mean M_i. Let the waiting time of the j^{th} member of the i^{th} group be X_{ij}, the sample mean waiting time for each group be $\overline{X}_i$, and the sample variance be S_i^2. Let the grand mean over all the groups be M_o and its sample value be $\overline{X}_o$.

By definition:

$$R^2 = 1 - \left[\sum_{i=1}^{G} \sum_{j=1}^{N_i} (X_{ij} - \overline{X}_i)^2 / \sum_{i=1}^{G} \sum_{j=1}^{N_i} (X_{ij} - \overline{X}_o)^2 \right] . \qquad (1)$$

Using the identity

$$X_{ij} - \overline{X}_o = (X_{ij} - \overline{X}_i) + (\overline{X}_i - \overline{X}_o) \qquad (2)$$

and the definition

$$S_i^2 = \sum_{j=1}^{N_i} (X_{ij} - \overline{X}_i)^2 / (N_i - 1), \qquad (3)$$

we can rewrite (1) as

$$R^2 = \left[1 + \sum_{i=1}^{G} (N_i - 1) S_i^2 / \sum_{i=1}^{G} N_i (\overline{X}_i - \overline{X}_o)^2 \right]^{-1} . \qquad (4)$$

Finally, since the standard deviation and the mean are nearly equal for a geometric distribution and since for large groups the sample mean will be close to the true mean, we can use as a reasonable approximation

$$R^2 = \left[1 + \sum_{i=1}^{G} N_i M_i^2 / \sum_{i=1}^{G} N_i (M_i - M_o)^2 \right]^{-1} . \qquad (5)$$

For instance, using the values of N_i and the estimates of M_i and M_o found in Table 2 of the text, we would expect regression analysis of a five-group mixture of geometric distributions to yield an R^2 value of approximately 0.15 at most.

MEDICAL CARE
February 1982, Vol. XX, No. 2

EPSDT Impact on Health Status: Estimates Based on Secondary Analysis of Administratively Generated Data

PATRICK H. IRWIN, PH.D. AND ROSEMARY CONROY-HUGHES, M.A.

EPSDT (early and periodic screening, diagnosis and treatment), a large-scale operational screening program that has generated a tremendous volume of data on the sociodemographic characteristics and health status of Medicaid-eligible children, seems to provide an ideal context within which to evaluate the effectiveness of preventive child care. Concern about health care expenditures generally, and the effectiveness of preventive child health services specifically, lead to the question of whether the impact on the health status of the children served can be measured without significantly adding to the cost of these services with primary data collection. We employ a quasi-experimental research design, using administratively generated data from an operational EPSDT program to estimate program impact on the prevalence of serious abnormalities among the children served. We found that, compared either with themselves across time or with a control group, a representative sample of 1,831 children had almost 30 per cent fewer abnormalities requiring care on rescreening.

Description of the Project

Background

RESEARCHERS in the field of child health have been unable to agree on the value of preventive services for children. However, in a recent federally sponsored critique of 38 controlled empirical studies, one reviewer argues that the credibility of the evidence on preventive care is significantly diminished by the poor methodological quality of much of the research. Although the studies are not totally inconclusive, the report concludes that "preventive child health care has, in general, been given only the weakest of tests."[1]

Evaluation of the long-term effectiveness of preventive care programs is a singularly difficult task, involving the costly and data-intensive process of measuring change over time. The increasing burden of health care expenditures suggests that such evaluation is essential to the development of cost-effective services. The costs of mass screening programs and other early-intervention measures can only be justified if those programs produce clear and lasting benefits. It is therefore imperative that evaluative evidence on the efficacy of preventive care be compiled and analyzed.

EPSDT (early and periodic screening, diagnosis and treatment), a large-scale operational screening program that has generated a tremendous volume of data on the

---

From the Philadelphia Health Management Corporation, Philadelphia, Pennsylvania.

This project was funded under Grant Number 18-P-97115, Health Care Financing Administration.

Address for reprints: Patrick H. Irwin, Ph.D., Philadelphia Health Management Corporation, 841 Chestnut St., Suite 1290.

0025-7079/82/0200/0216/$01.45 © J. B. Lippincott Co.

sociodemographic characteristics and health status of Medicaid-eligible children, seems to provide an ideal context within which to evaluate the effectiveness of preventive child health care. This study summarizes a methodology developed in southeastern Pennsylvania for estimating the impact of EPSDT; the results of this secondary data analysis, which employs a quasi-experimental research design, will undoubtedly be of interest to advocates and critics of preventive health services.

EPSDT in Southeastern Pennsylvania

EPSDT is a multiphase program of preventive health care for children of low-income families. The immediate goal of the program is to detect potentially debilitating health problems and, ultimately, to improve the health status and decrease the overall dependency of the target population. The screening procedure is *not* intended to be a complete diagnostic examination; rather, the screen serves to identify potential health problems and to distinguish those conditions that warrant immediate care or referral from those requiring no more than a simple follow-up. In Pennsylvania, all Medicaid-eligible children are entitled to be screened every three months for the first 18 months of life and every other year thereafter, through the age of 20.

The EPSDT program in the five-county Philadelphia area is administered by a private organization under contract to the State Department of Public Welfare and consists of five phases:

> I. Outreach or case finding, to identify and contact eligibles, to inform them of the program and to assist them in gaining access to services;
>
> II. Administration and coordination of screening, to establish and direct a network of qualified screening units to serve the area's target population;
>
> III. The actual testing or screening by certified providers of medical care;
>
> IV. Compilation and reporting of screening results; and
>
> V. Follow-through and diagnosis, to secure provision of indicated treatment services when necessary.

Over the past seven years of its operation, a management information system (MIS) has been developed, and large quantities of managerial data have been produced. By 1980, over 700,000 screens had been perfomed in this system and the medical records stored in a large computer file.

However, these administratively generated data are not ideally suited for research and evaluation, and they suffer from serious problems of data quality and completeness. The eligibility status of the target population changes rapidly, and there is an unavoidable time lag before a change in eligibility is posted to the data system; the program's MIS therefore does not provide an accurate estimate of the current number of eligibles at any point in time. Although screening data tend to be more complete than eligibility data, even this most reliable component of the MIS contains numerous instances of misidentification, duplicate records and anomalous data values. Finally since a relatively small proportion of children is screened more than once, longitudinal data are available for a very limited number of children.

Objectives of the Present Study

This study was undertaken as part of a larger effort to examine the process, effectiveness and costs of the EPSDT program in southeastern Pennsylvania, and it represents an attempt to determine the extent to which EPSDT has had an ameliorative impact on the health status of children receiving periodic screens (i.e., the program's effectiveness). As a demonstration, the project attempted to show to what extent, and how, the administratively generated data base could be exploited to estimate that effectiveness.

The project's initial attempts to measure the impact of the program on health status using only available data seemed to give ambiguous results. In terms of change in risk status between screen and rescreen (i.e., in the proportion of children having a serious problem on first screen compared with their second screen), no overall change is apparent in a sample of children who received two screens approximately two years apart (See Table 1).

The McNemar test results in a nonsignificant chi-square of 400/783 = 0.51 (df = 1, p = 0.48) and indicates that no *net* change in risk status is evident in this sample. Similarly, there is no evidence of significant changes in risk status whether or not the child returns to the same provider (nearly 58 per cent did) for rescreen. (See Table 2; chi-square is 0.037 and 1.786, respectively, with p > 0.10 for both.)

Tables 1 and 2 illustrate the impact of EPSDT by focusing on turnover—the clients who change risk status, from having a problem to not having one, and vice versa. In Table 1, 783 (402 + 381) children (42.7 per cent of the sample) changed status, even though the observed *net effect* over time is only approximately a 2 per cent decrease in those with problems. (Again, since it is not significant, the change in the population could be zero or even in the opposite direction for the entire population.) We may therefore conclude that there is no evident impact on health status.

From another perspective, in terms of the persistence of specific problems, the curative impact can be quite clearly estimated with the same sample. Of the 992 children with problems that require care at time 1, 40.5 per cent are problem-free at time 2. This improvement is probably *under*estimated, since we do not know from Table 1 whether the 590 children with problems at both times still suffered from the same abnormality identified at time 1 (nor the same number of abnormalities).

Deriving a reliable estimate of risk reduction, to determine if the same abnormality were present at both times, requires considerable data manipulation and assumptions about their meaning. Short of a medical record audit, we are able to determine whether the problems identified on rescreening were in the same general diagnostic area as those identified at the initial screen; if the subsequent problems are not in the same area, then they can be reasonably assumed not to be the same problem. Table 3 summarizes the results.

TABLE 1. Change in Risk Status from Initial Screen to Rescreen

		Time 2	
		No Problem (0)	Problem (1)
Time 1	(0)	450	381
	(1)	402	590

For the sake of presentation, we distinguish between risk reduction and problem resolution. A child is classified as "high risk" if there is at least one problem requiring care identified on screening. It can be argued, however, that risk is reduced if at least one of those problems is resolved. Surely some children classified as high risk at both times will have enjoyed reduced

TABLE 2. Change in Risk Status from Initial Screen to Rescreen for Same Provider and Different Providers

		Same Provider	
		Rescreen	
		No Problem (0)	Problem (1)
Initial screen	(0)	255	219
	(1)	214	366
		Different Provider	
		Rescreen	
		No Problem (0)	Problem (1)
Initial screen	(0)	203	162
	(1)	188	224

risk before rescreening but will have contracted another or a similar problem. In Table 3, a problem that "persists" (32.3 per cent of the total) may in fact not be the same abnormality but rather another problem in the same diagnostic area—another source of impact underestimation.

Undoubtedly, some problems requiring care will persist because the conditions are chronic. The remaining conditions that go untreated may be few indeed, which leads to a tentative conclusion that in terms of potential *curative* impact, the results are consistent with a significant improvement in health status among the EPSDT children in this sample.

While these conclusions from different approaches to the same data are not completely contradictory, they are ambiguous about the impact of EPSDT on health status. We attempt to resolve this ambiguity by developing a more complex research design, which incorporates longitudinal as well as cross-sectional comparisons, yet does not necessitate primary data collection.

Research Design

Conceptual Definition of Impact

Despite the limitations of the data system, it should be possible to determine, using only data generated by the program itself, the impact of EPSDT on the health status of the children it has served. Numerous confounding factors make this determination complex; nonetheless, it is important to know to what extent the impact can be isolated, without resorting to primary data collection or to experimentation with human subjects. "Impact" in this instance refers to a change in health status attributable to a given process (viz. EPSDT) and must be distinguished from those program performance measures that evaluate cross-sectional differences between groups rather than longitudinal change.

Changes in health status can be determined only for a specified individual or group passing through a given process. Differences, on the other hand, can be observed between or among any units of comparison. It is important to maintain the distinction between longitudinal changes and cross-sectional differences in this enterprise, as elsewhere.[2] The central methodological thesis of this study is that to analyze the impact of EPSDT on the health status of clients without collecting new data, it is necessary both to compare initial screen results with rescreen results *for the same children* and to compare children who have participated in the program with those who have not.

TABLE 3. Health Status at Time 2 of Children Who at Time 1 Had Had Problems Requiring Care (By Diagnostic Area)

Diagnostic Area	Problem Resolved		Problem Persisted		Total Problems
	n	%	n	%	
Dental	308	70.3	130	29.7	438
Vision	189	53.5	164	46.5	353
Hearing/speech	52	80.0	13	20.0	65
Development/behavior	41	85.4	7	14.6	48
Nutrition/growth	24	85.7	4	14.3	28
Medical	332	69.7	144	30.3	476
Total Problems	946	67.2	462	32.3	1408*

* Sums to more than the total number of problem screens (992) because 19.3 per cent of the children had multiple problems.

Central to this inquiry is a sample of children who received two screens approximately two years apart. Even if a simple comparison of health status (to be operationally defined below) at both times were to reveal improvements for this group of children, the differences could be due to something other than the benefits of the EPSDT assessment and referral. These rival explanations for observed differences in health status challenge our ability to attribute observed effects to the EPSDT program, but the challenge is a serious one *only if* the alternative explanations are plausible. The plausibility of competing explanations will be tested by comparing the two-screen sample with itself across time and with a control group of children just entering the program. To the extent that these comparisons tend to disconfirm any rival explanations, inferences about program impact are valid.

Comparative Design

Of the many assumptions that are necessary in this analysis, the most heroic is that impact—any observed change in health status—is due only to EPSDT. This admission is notably unsurprising. Outside (or inside!) a laboratory, unequivocal attribution of cause and effect is not a simple task. Still, if EPSDT as a health services program has been ameliorative, then we can reasonably expect some evidence of the effect. It is also reasonable that the possible sources of bias in available data be considered explicitly.

In any research endeavor, success in attributing observed (and hypothesized) change to some experimental treatment (e.g., the impact of a program) is primarily a question of the internal validity of the study. (The ensuing treatment of validity follows that of Campbell and Stanley.[3]) Did the program produce the change, or is there some alternative explanation (referred to as a "threat to the internal validity")?

We propose to examine explicitly the threats to internal validity that challenge this impact analysis. We begin, therefore, by repeating the eight primary threats to internal validity identified by Campbell and Stanley.

1. History—a specific event occurring between first and second measurement besides the experimental variable;
2. Maturation—changes occurring in the respondents as a function of time;
3. Testing—the effects of taking a test upon the scores of a second testing;
4. Instrumentation—changes in the observers or scorers producing changes in the obtained measurements;
5. Statistical regression to the mean—operating where groups have been selected on the basis of their extreme scores;
6. Selection—biases resulting from differential selection of respondents for experimental and comparison groups;
7. Experimental mortality—differential loss of respondents for the comparison groups;
8. Selection-maturation—and similar interaction effects.

Each of these sources of bias seems more or less plausible in the context of EPSDT impact analysis, with one significant exception. Testing cannot be considered as a threat to the internal validity of a research design in which the measurement instrument (the EPSDT screen) *is* the experimental treatment.

Random assignment of subjects to experimental and control conditions may be the optimal design with which to assess the impact of social programs, but this decisive, single-experiment approach is not feasible in most field settings.[4] Without the use of randomization, quasi-experimental research designs can only partially resolve the issues of threats to internal validity. Fortunately, in contrast to the conduct of research in an experimental environment, research conducted in situ offers a high potential for findings that can be readily applied or generalized to other operating environments (i.e., it offers high external validity).

The potential for application across contexts is crucial in situations in which inquiry will always be subject to challenges to internal validity. If the same impact is found in several contexts or in programs that are each subject to different uncontrolled variable effects, then the evidence for the single explanation of program impact grows "just because the plausible rival hypothesis is different from study to study."[3] This inferential feature is deliberately introduced in the following design. The design also attempts to provide not only a way around the ethical impasse to human experimentation by denial of indicated treatment, but also to demonstrate a cost-effective and unobstrusive mode of empirical research in preventive child health services.

Comparison Groups

Overall, the research design is comparative, with observations made both cross-sectionally (to measure differences) and longitudinally (to estimate changes). The groups to be compared are a two-screen sample (the experimental group), and two other samples of initial screens approximately two years apart.

Using the Campbell and Stanley diagrammatic convention, the research model may be formulated as follows:

Two-Screen Experimental Group: O1 X O2
One-Screen Control Group (1977): O3
One-Screen Control Group (1979): O4

The "0" stands for an observation at a given time, "X" for experimental treatment. Technically, in this case, the observation is the EPSDT health screen, and the treatment is the referral. The EPSDT assessment and referral are separable only conceptually, however, since our concern is the outcome of exposure to an EPSDT screen. The comparison is between those who have had this exposure previously and those who have not.

Sample of Rescreened Children. For a child to be included in the two-screen sample for this analysis, several criteria have to be met: 1) the child's first screen occurs between July 1, 1977, and December 31, 1977; 2) the child is at least 19 months of age at the time of the initial screen; and 3) the time interval from first to second screen is at least 18 (and up to 32) months. (After the restrictions to the sampling frame were imposed, only 1,831 children qualified, so all were included in the two-screen group. Thus, these children do not constitute a simple random sample, although this virtual enumeration is no less representative than a random sample.)

Criterion 1 insures first that these children all represent first-to-second rescreens rather than second-to-third, etc. Using a more homogeneous sample should improve the precision of our estimates of risk reduction. The beginning of the six-month time frame selected for the first screen corresponds to an internal change in the data collection system (in July 1977) that changed certain data items and codes.

Criterion 2 insures that we have not selected children who are 18 months of age or less and consequently on a special infant rescreening schedule. These children receive a somewhat different series of tests and are permitted to receive six screens from birth through the age of 18 months. Our interest in this exercise is focused on children who are eligible for the two-year schedule.

Criterion 3 insures that the two-year schedule has been followed at least approximately. Again, this should help to provide a more homogeneous sample.

Control Groups. The first control group (the 03 sample of the preceding diagram) consists of a sample of all children receiving their first screen at about the same time as the two-screen sample in August 1977, and who were also at least 19 months old (n= 2889). This group overlaps with children in the 01–02 sample—those of the two-screen group who received their initial screen in August 1977. Some of the comparisons to be made assume the independence of samples and require the exclusion of this overlap; we shall refer to the restricted 03 sample as 03′, n = 2,578. (Again, for both control groups, limitations to the sampling frame, cost and convenience led us to include all who qualified in these nevertheless representative groups.)

The second control group (04) consists of children in the same age range who receive their first screen at approximately the same time that the two-screen sample was being rescreened, between January and April 1979. An effort was made to further restrict this group to those children from families who were new to the program as of January 1979, to insure sample independence (n= 1,182).

Operational Definition of Impact

A complete EPSDT screen in Pennsylvania includes as many as 39 individual Test Areas (TAs), organized her into six mutually exclusive summary Diagnostic Areas (DAs): medical, vision, hearing, dental, laboratory and behavior. (This study does not explicitly consider assessment of the self-reported immunization status, which is also part of the screen.) Some of these TAs may not be assessed in every screen because they are discretionary or are not appropriate for a given child.

The response categories are not identical for all TAs, but the alternatives provided for the medical evaluations are typical:

Code	*Response*
1	Normal
2	Abnormal, treatment required
3	Abnormal, nontreatable
4	Not done

For the purpose of this analysis, only code 2 (abnormal treatment required) is taken as evidence of a "serious" problem. Whether the child is referred or not is apparently more indicative of a provider's willingness, specialization or experience in treating the condition rather than a medical judgment or assessment of the child's health status. Codes 3 (abnormal, nontreatable) and 4 (not assessed) are counted as missing values.

For the longitudinal and cross-sectional comparisons required in the present study, we operationally define Health Status (HS) in terms of the total number of abnormalities that require treatment. Again, abnormalities that are identified as not requiring care or as not treatable (a relatively rare occurrence) are excluded. To standardize HS for comparisons among different groups of children, the total number of Abnormal Test Areas (ATAs) found is divided by the total number of test areas that are assessed for a group of children. A low score on this index therefore indicates a healthy individual or group; a higher score indicates a relatively higher incidence of illness or injury.

The measure of HS is calculated according to the following formula and can be expressed either as a proportion or a percentage:

$$\mathrm{HS} = \frac{\mathrm{ATA}}{\mathrm{TTA} - \mathrm{NA}}$$

Where, for a specified group of children, HS = health status index; ATA = total number of *abnormal* test areas where treatment is required; TTA = TAs × S, the number of test areas in which a treatable abnormality can be found (39) times the number of children screened; and NA = total number of test areas not assessed, an adjustment for a given group of children to eliminate TAs not assessed.

With a slight modification of this formula, we can also confine the focus of the measure to a particular DA; i.e., any specified selection of TAs. For example, if we wish to consider only conditions within the dental DA, then the potential number of TAs is limited to three possible tests: caries, oral infections and other dental problems. When a particular subset of TAs (< 39) is selected for calculating the index, we shall refer to it as an Abnormality Rate (AR).

The HS index can be interpreted as the probability that the EPSDT screen will uncover any condition requiring treatment. Similarly, an AR for a specified DA can be interpreted as the probability that a specific kind of problem will be identified.

Adjustment Estimation Procedure

Before controlling for the relavant threats to internal validity, there seems to be virtually no change in HS for the longitudinal sample between screen and rescreen, and no difference in HS for the cross-sectional samples between rescreen and control group. For the unadjusted longitudinal sample, 2.61 per cent of all TAs assessed on initial screening in 1977 were abnormal, with a condition requiring treatment; that percentage appears to have *changed* by 1979 to 2.69 per cent on rescreening. This latter health status level is also compared with 2.58 per cent for the control group of children who received their initial screen in 1979.

To conclude at this point that there is evidently no impact would be premature. Particularly in light of the need to control for rival alternative explanations, we maintain that these unadjusted comparisons are invalid: the groups are not yet comparable.

In the sections that follow, we describe in some detail the means of estimating the needed adjustments and then compare the groups again after adjustment; actual adjustments are calculated in the "Findings" section. The procedures for estimating and adjusting may seem complex, but the conclusions about EPSDT impact on health status that we can make at that point will differ significantly from preceding conclusions based on unadjusted HS. While not all sources of incomparability of groups can be controlled in each comparison made, by integrating two modes of comparative analysis—longitudinal comparison of change with cross-sectional comparison of difference—we can gain a new perspective that is essential for evaluating impact based on this administrative data base.

After estimating the magnitudes and directions of each of the adjustments needed, the adjustment procedure is 1) to ascertain the per cent of change or differences due to factors other than the program—the threats to internal validity—and then 2) to add or subtract this percentage to/from the per cent of difference or change in the unadjusted HS rates. For example, if there has been a 10 per cent increase in abnormalities due to instrument changes, we would subtract this figure from the per cent change in HS for the longitudinal comparison between screen and rescreen. The effects of this procedure will be tabulated both for the overall HS index and the selected DAs.

Instrument and History Effects. To gauge changes in health status over time, the first requisite is that the assessment (instrument) remain constant. A less rigorous procedure at rescreen than at initial screen, for example, would probably result in fewer abnormalities being found and could mislead us into accepting an apparently improved health status that is actually due to instrument change. When there are events extraneous to the EPSDT program, which clients commonly experience between screenings, these events could alter health status (history). The introduction between screens of a new child health initiative for Medicaid-eligible children (e.g., an intensified nutrition or immunization program within the public schools) could mislead us into attributing improvements in health status to EPSDT rather than to history.

Instrument or history effects on health status assessment are analogous to the effect of inflation on income: health status may appear to improve across time only because its measurement is less rigorous. As with inflation, it is possible to control statistically for differences due to changes in instrument or history, but only if they can be identified and isolated from changes attributable to the program.

A test of the instrument/history hypothesis can be made by comparing the two control samples of initial screen results—in terms of the diagrammatic convention, the 03–04 comparison. Since the 03 and 04 control groups are independent and representative samples of chidren being screened at two times, differences between them can be attributed to instrument or history effects. Those observed differences that are also confirmed by program history as resulting from protocol or programmatic changes can be used to estimate the magnitude of the effect and to adjust apparent changes in the rescreened (i.e., experimental) group.

Of the 39 individual comparisons included in the HS index, nine revealed differences in ARs that are significant ($p < 0.05$). In general, we found higher ARs in the later (04) sample; this finding suggests that the screening protocol may have become more rigorous over time.

EPSDT program administrators have offered plausible explanations for virtually all of the observed differences. Administrative personnel attribute the differences in ARs to changes in EPSDT procedures or to changes in the Medicaid service delivery system that have occurred in this area in the two years between the screens. Thus, examination of independent groups from time 1 and time 2 suggests that factors relating to instrument and history may indeed threaten the internal validity of our longitudinal comparisons. (See Table 4.)

Table 4 presents the HS index and selected ARs for the 1977 and 1979 cross-sectional samples (03 and 04, respectively). The overall HS in the 1979 sample is approximately 21 per cent higher than the corresponding rate in the 1977 sample. This estimate of instrument effect will be used below to adjust the per cent change in HS between 01 and 02. (ARs are adjusted similarly according to their respective 03–04 differences.)

Because the age distributions of the 03 and 04 samples are not the same, any AR differences between the groups may be attributable to age differences rather than to instrument changes. To age-adjust at this stage would introduce a new confound, but to avoid the possibility of confusing age differences with true instrument/history effects, we compared the ARs in two other

TABLE 4. Selected First Screen Abnormality Rates (ARs) in 1977 (03) and 1979 (04)

DA (# TAs included)		03		04		
		Total Tests Done	AR%	AR%	Total Tests Done	p-value
HS	(39)	83,608	2.13	2.58	34,407	0.00
Medical	(13)	37,542	1.30	1.96	15,375	0.00
Vision	(2)	4,508	9.05	9.96	1,969	0.26
Hearing	(2)	4,502	1.38	1.99	1,956	0.09
Dental	(3)	8,667	7.03	6.26	3,549	0.12
Laboratory	(10)	12,635	0.87	1.25	5,436	0.02
Behavior	(9)	15,754	0.67	1.00	6,122	0.02

Note: Indicated p-values are the two-tailed significance levels resulting from a difference-of-proportions test.

cross-sectional groups that, while not carefully selected samples of first screens, are already age-comparable. Using administrative statistical reports for July through December 1977 (covering 36,415 screens) and for October 1979 (4,424 screens), we found differences similar to each of those reported here.]

The estimated difference may reflect increasingly rigorous quality control and standardization of tests (e.g., "not assessed" vision TAs decreased from about 22 per cent in 1977 to 17 per cent in 1979). In addition to the HS adjustment, specific TAs vary in the amount and significance of difference. Therefore, any screen–rescreen change in HS or ARs (except Dental) toward improvement will be underestimated and therefore should be increased by the same amount. (Dental AR change will be overestimated because the probability of finding a dental abnormality decreased, although not significantly.)

Selection and Statistical Regression Effects. In gauging impact by cross-sectional comparison of rescreen results with the initial screen results of a control group, we realize that observed differences in health status could also be due to differential selection or statistical regression. If those who received two screenings represent another, perhaps healthier, population, selection bias or statistical regression to the mean may explain any cross-sectional differences.

A test of this rival hypothesis that is possible with the available data involves comparison of the longitudinal sample at initial screen with others being screened initially at that time; i.e., the 03′ control group. If those children who eventually received rescreening are significantly different (e.g., healthier) from others at the outset, a selection bias is indicated. The 01–03′ difference can be used to estimate the magnitude of the bias and to adjust the cross-sectional comparison of the rescreen results (02) and control group results (04). If the rescreened group enjoys the same health status as others being initially screened at that time, we know at least that the rescreened children did not have a head start.

Unfortunately, since there could be no random assignment from all children who were screened once to the group of children who were rescreened, rescreening may itself be taken as evidence that the children are generally different (i.e., extreme) in terms of health, specifically in tendency to improve, a selection–maturation interaction.[3]

Tendency to improve cannot be isolated easily using a longitudinal control group, observed both at time 1 and time 2 but not treated, because the health assessment (or experimental observation) is the "treatment." Still, by extending the comparison at time 1 of the initial screen children (01–03′) to nonhealth factors (e.g., demographic comparisons of available data), we might still support the claims that the rescreen group is equivalent to the control group at initial observation.

Receptiveness to intervention is a characteristic not necessarily discernable from initial HS; but some exogenous sociodemographic factors might present children with treatable abnormalities an environment that potentiates the effects of the EPSDT intervention. In other words, the chances that any indicated treatment regimen will be followed may be higher in certain households and among certain classes of children than in others.

Table 5 displays the summary DAs for 01 and 03′, including general HS. Generally, the two-screen sample has a significantly ($p < 0.0001$) higher abnormality rate at first screen (26.1 per cent higher) than does the control group. The medical, vision and dental DAs are the chief contributors to this difference.

Rather than enjoying a head start as compared with others receiving their initial screening at the same time, the lon-

TABLE 5. Selected First-Screen Abnormality Rates (ARs) in Two 1977 Samples

DA (# TAs included)		01		03′		
		Total Tests Done	AR%	AR%	Total Tests Done	p-value
HS	(39)	53,446	2.61	2.07	97,929	0.00
Medical	(13)	23,801	1.58	1.27	33,500	0.00
Vision	(2)	2,975	11.97	9.03	4,010	0.00
Hearing	(2)	2,975	1.68	1.30	3,997	0.20
Dental	(3)	5,493	8.47	6.66	7,734	0.00
Laboratory	(10)	8,392	0.94	0.84	11,279	0.46
Behavior	(9)	9,807	0.71	0.68	14,009	0.78

Note: 03′ consists of all August 1977 initial screens *excluding* those belonging to 01 sample. Indicated p-values are the two-tailed significance level resulting from a difference-in-proportions test.

gitudinal sample appears to have had proportionately more serious problems. There are several possible reasons for this selection effect. One is that the identification of risk increases the probability that a child will be rescreened (i.e., the presence of a treatable abnormality on initial screen co-occurs with an increased incidence of rescreening of almost 29 per cent). While almost 60 per cent of the 03′ group were known to be eligible for EPSDT through May 1979, there is no record that any of them was rescreened by March 1980. Based on the entire 03 sample of 2,889 children who were screened for the first time in August 1977, 1,428 had problems, and 1,461 did not. Of those with problems, 173 or 12.1 per cent were rescreened, compared with 138 or 9.4 per cent of those without problems. The association between whether or not a child has a problem and whether or not the child will be rescreened is significant (chi-square = 5.32, df = 1, $p < 0.05$).

Another possible reason for the higher 01 ARs is seasonality, or the fact that the longitudinal sample is drawn from a six-month sampling frame as compared with a one-month frame (August) for the comparison group. Other possible reasons for the 01–03′ difference may lie in the differences in their age distributions (chi-square = 82.9, df = 17, $p < 0.05$) and in differences in the distributions of the providers who screened them (chi-square = 323.05, df = 74, $p < 0.05$). (Both chi-square statistics for age and provider differences are calculated after excluding categories with expected frequencies of less than 5.)

In order to test the possibility of a selection–maturation interaction or tendency to improve, the two groups (01 and 03′) were also compared on a series of demographic variables. In terms of sex, race and whether the child is from Philadelphia or the suburban counties, there are no significant differences. In terms of payment category ("categorically needy" or "medically needy" *versus* "general assistance"), household size and age distributions, the differences are statistically significant ($p < 0.05$) but otherwise very small. (See Fig. 1A and 1B.) [Other socioeconomic comparisons can be made of measures derived from 1970 Census tract-level data for most of the children in the samples. Given the nature of the data, these comparisons are extemely tentative, but they may provide a useful sense of environmental factors. The average (mean) median family income differs significantly between groups ($6,337 for 01 and $6,606 for 03′, $p < 0.05$), while

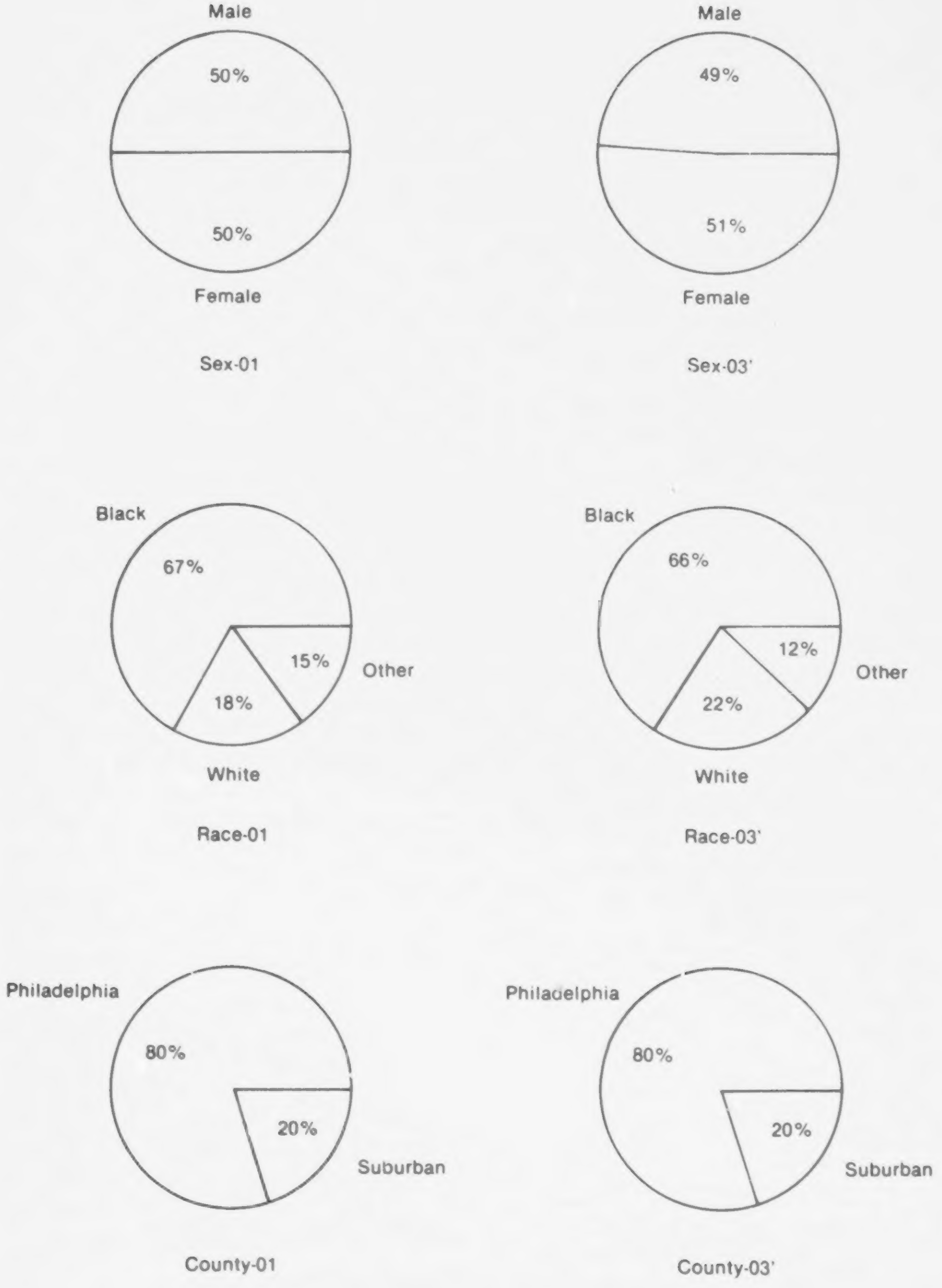


FIG. 1A. Demographic comparisons at first screen between those who received a rescreen (01) and those who did not (03').

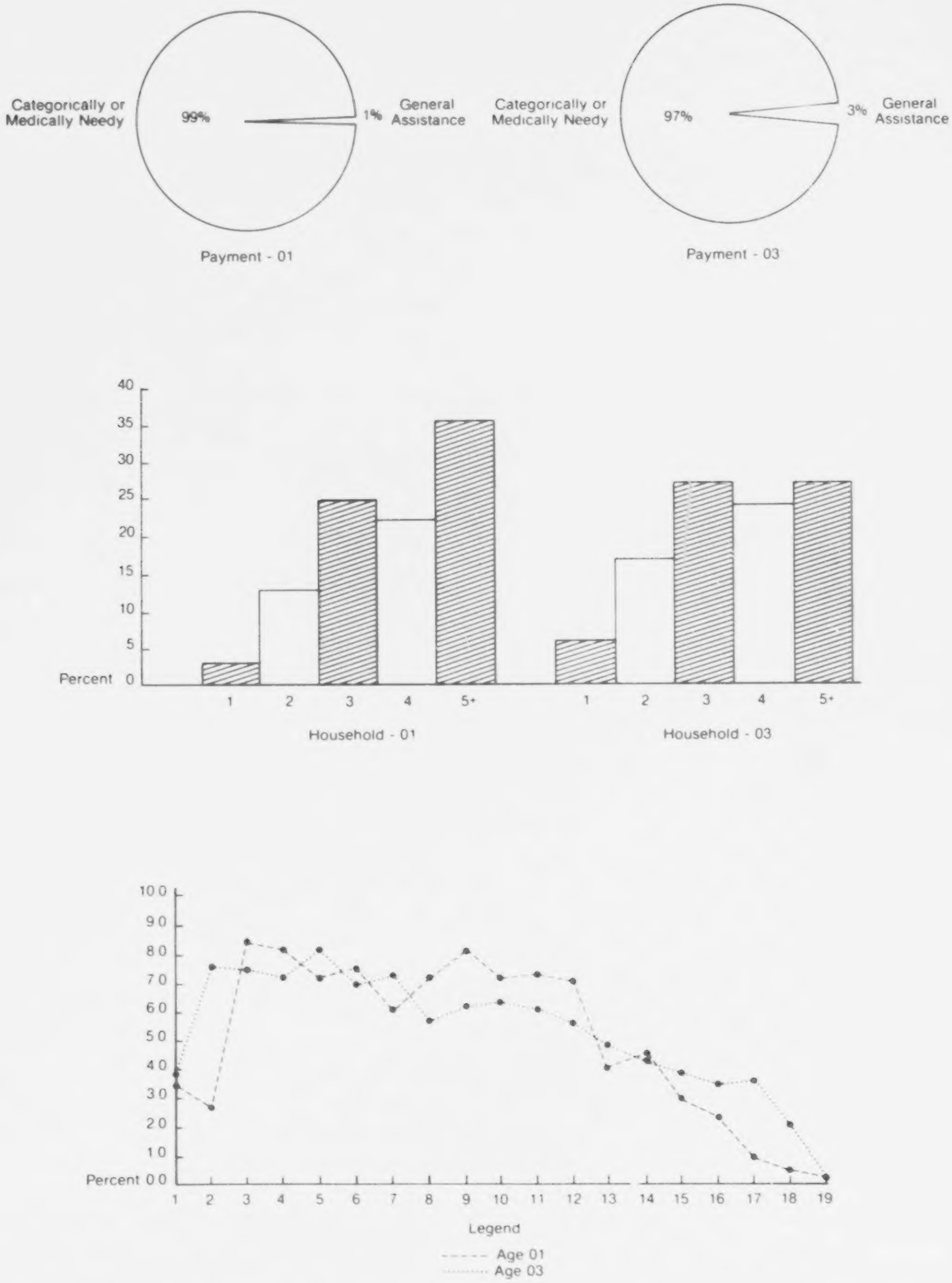


FIG. 1B. Demographic comparisons at first screen between those who received a rescreen (01) and those who did not (03').

the mean per cent living in crowded housing does not differ significantly (18.1 per cent and 19.8 per cent, respectively).]

Generally, the demographic differences described are interesting and suggestive, but, in terms of practical significance in support of the selection–maturation interaction hypothesis or as a source for an adjustment to the longitudinal comparison, they are too tentative. Moreover, we note that the coeval cross-sectional comparison does not suffer from this interaction threat to validity, as does the longitudinal comparison. If the results of both comparisons are the same, after other threats are controlled, tendency to improve would lack support as a rival hypothesis.

Maturation Effects. Another of the plausible alternative explanations for observed longitudinal changes is maturation. Campbell and Stanley[3] define maturation as "processes within the respondents operating as a function of time." As such, maturation is a form of change and can affect only the longitudinal (01–02) comparison. Inferences drawn from cross-sectional comparisons (02–04), on the other hand, may be invalidated by age differences between the experimental and control groups. An age-adjustment procedure must therefore be used to control for both maturation and age differences.

The mechanics of age adjusting involve a procedure analogous to deriving standardized specific rates (e.g., age-specific birth or death rate). The age standardization recalculates HS (or specific ARs) as if the two groups were identically composed on age. The formula is simply a weighted mean that expresses the age-specific HS of the two first-screen groups (01 and 04) weighted by the age distribution of the rescreened children (02).[5] If there are no age differences between any two comparison groups, then standardization will not change the HS, although the converse is not necessarily true.

Since no child in the rescreened group is under three years of age when they are rescreened, and none over 18 in the first-screen groups, it is important to note that standardization will limit the comparisons to only those children 3 through 18 years of age. (See Fig. 2A-C.)

Experimental Mortality. The validity of comparing rescreened children's health status with that of children just receiving their initial screens (02–04) is also threatened by the fact that, if the program's future emulates its past, most of the children in the comparison group (04) will not be rescreened. Since there are differences at first screen between those children who are eventually rescreened and those who are not (the selection effect), we may therefore suspect that, had all first-screen children received a second health assessment, our estimation of EPSDT impact would be different. The differential loss of children from the time series implies that the observed cross-sectional 02–04 differences may be due to "experimental mortality" rather than to EPSDT.

Longitudinal comparisons (01–02), on the other hand, do not suffer from the effects of differential selection, since the same group is compared intact across time. Therefore, although this research design requires both cross-sectional and longitudinal comparisons, only the inferences from cross-sectional data are subject to the challenge of experimental mortality.

The plausibility of experimental mortality as an alternate explanation for observed cross-sectional differences might be tested by locating and screening those children who were not yet rescreened. If this new estimate of health status for rescreened children does not differ from that of the previous rescreened group, the hypothesis of experimental mortality could not be supported. Even if the test were feasible, however, the resources required to locate and rescreen these children would be prohibitive.

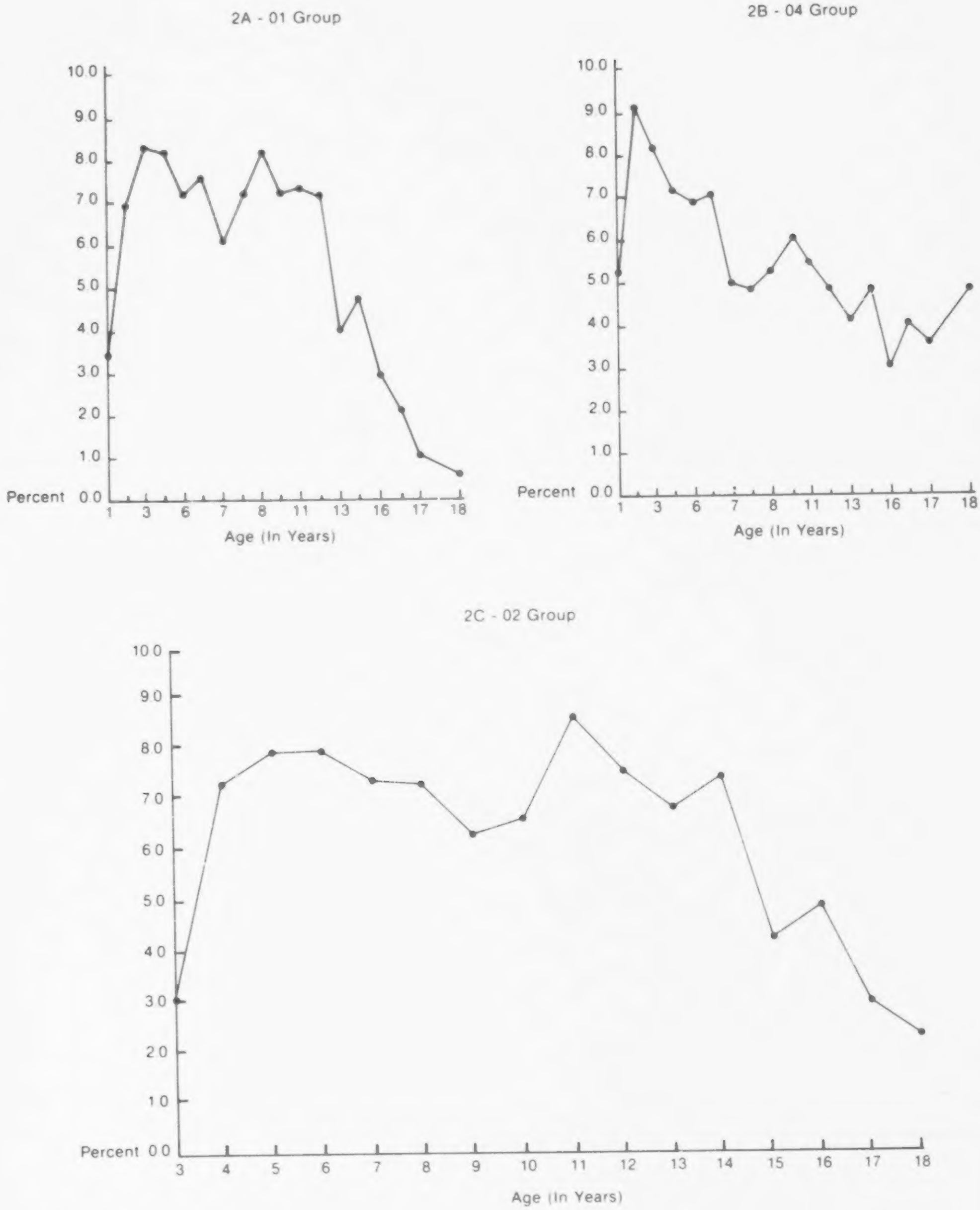


FIG. 2. Age distributions among the comparison groups. 2A, 01 group; 2B, 04 group; 2C, 02 group.

Given the practical difficulties of eliminating experimental mortality as a rival explanation for cross-sectional differences, it is appropriate to ask what, if any, effects this factor could possibly have on the estimation of impact in this study? The evidence suggests that most arguments that come to mind are in the wrong direction. For example, it might be assumed that those who drop out will be less healthy because they are likely to be less conscientious about their health and medical care than those who stay. In fact, as was seen in Table 5, the opposite is true: the experimental group is less healthy at initial screen than the control group.

There does not seem to be a feasible way to control for experimental mortality in the cross-sectional (02–04) comparison. We can, however, again use the inferential feature of this research design to combine the results of cross-sectional and longitudinal comparisons. If the longitudinal (01–02) comparison, which is unaffected by experimental mortality, discloses the same impact after all other threats have been controlled, then there would be no empirical evidence that the EPSDT program impact hypothesis is less plausible. To the extent that they differ, experimental mortality may be the source.

Findings

Age Adjustment. Table 6 shows fairly consistent increases in 01 and 04 ARs when the age adjustment is made. The difference between each group's crude and standardized health status must be attributed to the differences between age distributions compared with the two-screen group (02) at rescreening; for the 01 group, there is about 8 per cent difference between its unstandardized and standardized HS (2.61 per cent versus 2.83 per cent) and 6 per cent difference for 04 (2.58 versus 2.74 per cent). In both comparisons, the effect of age standardization is to improve the experimental (02) group's health status relative to those of the comparison groups. In other words, compared with children at rescreening (02), their own health status at first screen (01) and that of a control group being screened for the first time (04), they appear to be slightly worse due (first) to the favorable age distributions of the first-screen groups.

Longitudinal Adjustment. After the age distributions of the control groups have been standardized, further adjustments are needed: observed changes (01*–02) must be adjusted to control for the effects of increased screening rigor, and cross-sectional differences (02–04*) must be ad-

TABLE 6. Comparison of Abnormality Rates for 02 with 01 and 04 Before and After Age-Adjustment

DA	AR%						
	02	01		01*	04		04*
HS	2.66	2.61	→	2.83	2.58	→	2.74
Medical	1.98	1.58	→	1.55	1.96	→	1.82
Vision	11.95	12.00	→	13.13	9.96	→	11.01
Hearing	2.15	1.68	→	1.79	1.99	→	2.09
Dental	7.13	7.57	→	9.79	6.26	→	7.58
Laboratroy	1.06	0.94	→	0.85	1.25	→	1.08
Behavior	0.63	0.71	→	0.70	1.00	→	1.05

* An age-adjusted rate.

justed to control for the effects of self-selection for rescreening. Differences that are observed in both the longitudinal and cross-sectional comparisons of adjusted abnormality rates may then be interpreted as evidence of program impact.

The adjustment procedure applied to instrument changes is simply to calculate the per cent change from 01* to 02 and then to subtract from that the per cent change from 03 to 04, due to instrument/history change. The adjustment factors are based on estimates from Table 4 and are signed positive if more abnormalities are found in 1979 than in 1977, and negative if fewer abnormalities are found (Table 7). Similarly, the signs of the unadjusted and adjusted per cent changes indicate the direction of change in ARs before and after adjusting for instrument effect—a negative change indicates reduced abnormalities.

Before the adjustment for instrument effect, there is evidence (Table 6) to support claims of some improvement in health status between age-comparable screen and rescreen results across time. When we adjusted for the increased rigor of the 1979 screening protocol, compared with that in 1977, we find the abnormality rate in 1979 to be about 27 per cent lower. Thus, in terms of the original HS index for the two-screen group in 1977, an average of 26.1 abnormalities were found per 1,000 tests; for the same children in 1979, 19.0 abnormalities per 1,000 tests were found.

Table 7 also allows us to estimate relative impact in different diagnostic areas, although it should be kept in mind that these estimates are based on differing numbers of tests and are therefore of variable reliability. Even after adjusting for age and instrument changes, the greatest improvement appears in the behavioral area, with an almost 60 per cent lower AR.

Cross-Sectional Adjustments. Since cross-sectional comparisons of rescreen results (02) with those of an age-adjusted control group (04*) are subject to selection/regression effects, the per cent differences in rates between 02 and 04* are adjusted in Table 8 based on estimates from the 01–03′ comparisons in Table 5. The sign of the adjustment factor is positive for all DAs, indicating that the two-screen sample had on average higher abnormality rates on first screen than others receiving first screens at the same time. At this point we must recognize that having a significantly sicker group of children for the two-screen sample (at time 1) could mean that the ob-

---

* Age distributions of control groups have been standardized.

TABLE 7. Longitudinal Changes Adjusted for Instrument/History Effects

	Per Cent		
DA	01*–02 Change	03–04 Adjustment Factor	Adjusted 01*–02 Change
HS	− 6.01	+21.13†	−27.14
Medical	+27.74	+50.77†	−23.03
Vision	− 8.99	+10.06	−19.05
Hearing	+20.11	+44.20	−24.09
Dental	−27.17	−10.95	−16.22
Laboratory	+24.71	+43.68†	−18.97
Behavior	−10.00	+49.25†	−59.25

* An age-adjusted rate.
† Adjustment factors based on significant 03–04 differences ($p < 0.05$). (See Table 4.)

served differences between 02 and 04* are at least partly due to statistical regression in spite of the control. Campbell and Stanley suggest, however, that in the case in which the experimental group is self-selected—such as the two-screen sample—the problem of uniform regression effects between experimental and control group becomes less likely, but interaction effects are more likely.[3]

The adjustment procedure summarized in Table 8 parallels that used for the longitudinal comparison. In this case the per cent AR difference between 01 and 03′ is subtracted from the per cent difference between 04* and 02. As in the longitudinal comparison, the negative signs of several 04*–02 differences, including the overall HS, indicate that the rescreened children are slightly healthier than those of similar age just entering the system in 1979. The estimated adjustment for selection/regression results in the rescreened group showing about 29 per cent fewer abnormalities than the age-adjusted comparison group at first screen.

Conclusions. The similarity of results of the two modes of comparison is more that accidental: the rescreened group has between 27 and 29 per cent lower overall abnormality rate, whether compared with itself over time or with a control group.

It is still possible for either one of these differences in isolation to be explained by rival hypotheses. The longitudinal difference could be due to uncontrolled statistical regression or selection–maturation (and other) interaction effects; the cross-sectional difference may be attributable to experimental mortality. But the plausible rival hypotheses are different in the two comparisons. The research design included this inferential feature of changing contexts deliberately to enable us to isolate and gauge the impact of the EPSDT program based on secondary data analysis.

Summary

The findings presented here attest to the beneficial effect of EPSDT on the health status of children served. Although narrowly focused, our conclusion is definitive and unequivocal: periodic screening is associated with a decrease in the prevalence of abnormalities requiring care.

The results of this study do not permit a definitive statement on the ultimate impact

---

* Age distributions of control groups have been standardized.

TABLE 8. Cross-Sectional Differences Adjusted for Selection/Regression Effects

	Per Cent		
DA	04*–02 Difference	01–03′ Adjustment Factor	Adjusted 04*–03 Difference
HS	− 2.92	+26.09†	−29.01
Medical	+ 8.79	+24.41†	−15.62
Vision	+ 8.54	+32.56†	−24.02
Hearing	+ 2.87	+29.23	−26.36
Dental	− 5.94	+27.18†	−33.12
	− 1.85	+11.90	−13.75
Behavior	−40.00	+ 4.41	−44.41

* An age-adjusted rate.

† Adjustment factors based on significant 01–03′ differences ($p < 0.05$). (See Table 5.)

of EPSDT and certainly cannot be construed as a blanket endorsement of screening programs in general. We have, on the contrary, deliberately restricted the operational definition of "impact." Our concern is with health status impact, and not with such equally important outcome measures as equity, cost–benefit or client satisfaction. Furthermore, the term "health status" is used exclusively in terms of abnormality rates and does not measure level of functioning, risk status or any of the other significant dimensions of well-being. Health status (and, ultimately, impact) can be measured only by isolating a set of very specific variables on which data are available and from which valid conclusions may be drawn. Consequently, the results of this analysis cannot be interpreted as an estimate of the only, or even of the most important, impact of the EPSDT program. Also, these results do not permit an evaluation of the extent to which EPSDT has achieved its ultimate goal of decreasing overall dependency in the target population. In fact, such an evaluation may never be possible for a program with ultimate objectives as complex and comprehensive as those of EPSDT. Recognition of these limitations in no way diminishes the significance of specific impact that has been identified and documented.

The credibility of any impact estimate depends to a great extent on the power of the research design to eliminate competing explanations for the observed phenomena. Thus, the distinctive methodology employed in this study suggests policy implications that are no less important than those suggested by the substantive conclusions. Specifically, the value and feasibility of secondary data analysis as a method for evaluating program impact has been convincingly demonstrated.

However, it is important to recognize that such secondary analysis is not easily accomplished. Our initial attempts to measure impact produced ambiguous results—no significant change in risk status occurred between first screen and rescreen, despite clear evidence of problem resolution. The actual changes in health status were revealed only after a complex series of mathematical adjustments.

In spite of its complexity, the research design employed here offers several distinct advantages. Using a quasi-experimental design, we can estimate impact without incurring the cost of primary data collection; in addition, such a design avoids the ethical difficulties of selective denial of treatment. Because it is both cost-effective and unobtrusive, the methodology used in this study may be applicable to other operating environments. Through the use of similar secondary data analysis, a national EPSDT monitoring system could be developed, and the impact of screening on the health status of children in a variety of environments could be estimated.

EPSDT, like other public service programs, presently finds itself in an environment of both scarce and shrinking resources and a constant or increasing demand. Estimating the impact of such a program on those it serves is therefore an essential first step toward improving the efficiency of that program by maintaining or increasing the level of services but at a lower cost.

References

1. Shadish WR. Effectiveness of preventive child health care. HCFA Grants and Contracts Reports publication no. 03099, April 1981.

2. Lieberson S, Hansen LK. National development, mother tongue diversity, and the comparative study of nations. American Sociological Review 1974;39:523.

3. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1963.

4. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand-McNally, 1979.

5. Mueller JH, Schuessler KF, Costner HL. Statistical reasoning in sociology. Boston: Houghton-Mifflin Company, 1977.

CS170-1

Medical Care

March 1982
Vol. 20 No. 3
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Evaluation Methodology for Ambulatory Care Information Systems

How Good Are the Data? Reliability of One Health Care Data Bank

The Interaction Among the Criteria Physicians Use When Prescribing

Does Cost Information Availability Reduce Physician Test Usage? A Randomized Clinical Trial With Unexpected Findings

The Effects of Price Information on Physicians' Test-Ordering Behavior: Ordering of Diagnostic Tests

Physician Sensitivity to the Price of Diagnostic Tests: A U.S.–Canadian Analysis

Why Veterans Choose Veterans Administration Hospitalization: A Multivariate Model

Satisfaction With Maternity Care: A Matter of Communication and Choice

Utilization of Specialized Ambulatory Care by the Elderly: A Study of a Clinic

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year for individuals, $45.00, for institutions, $50.00; special price U.S. students, $27.00; all other countries except Japan, one year, for individuals, $55.00, for institutions, $60.00. Single copies $6.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

March 1982
Volume 20, Number 3

CONTENTS

POLICY FORUM

Medically Oriented Computer-Based Information Systems. *Alan F. Dowling, Jr.* 253

ORIGINAL ARTICLES

Evaluation Methodology for Ambulatory Care Information Systems. *Donald W. Simborg and Quinn E. Whiting-O'Keefe* 255

How Good are the Data? Reliability of One Health Care Data Bank. *Leslie L. Roos, Jr., Noralou P. Roos, Sandra M. Cageorge and J. Patrick Nicol* 266

The Interaction Among the Criteria Physicians Use When Prescribing. *Robert N. Zelnio* 277

Does Cost Information Availability Reduce Physician Test Usage? A Randomized Clinical Trial With Unexpected Findings. *David I. Cohen, Paul Jones, Benjamin Littenberg and Duncan Neuhauser* 286

The Effects of Price Information on Physicians' Test-Ordering Behavior: Ordering of Diagnostic Tests. *K. Michael Cummings, Kenneth B. Frisof, Michael J. Long and George Hrynkiewich* 293

Physician Sensitivity to the Price of Diagnostic Tests: A U.S.—Canadian Analysis. *John Hoey, John M. Eisenberg, Walter O. Spitzer and Duncan Thomas* 302

Why Veterans Choose Veterans Administration Hospitalization: A Multivariate Model. *William Frank Page* 308

Satisfaction With Maternity Care: A Matter of Communication and Choice. *Deborah A. Sullivan and Ruth Beeman* 321

Utilization of Specialized Ambulatory Care by the Elderly: A Study of a Clinic. *Stanley R. Ingman and Ian R. Lawson* 331

BOOK REVIEWS

Medical Records in Health Information, by Kathleen A. Waters and Gretchen Frederick Murphy. *Faye Brown* 339

Managing Health Records: Administrative Principles, by Joan Gratto Liebler. *Faye Brown* 339

Computers in Nursing, edited by Rita D. Zielstorff. *Barbara J. Horn* 340

Nursing Information Systems, edited by Harriet H. Werley and Margaret R. Grier. *Barbara J. Horn* 340

Improving Doctor Performance: A Study in the Use of Information and Organizational Change, by Steven Kelman. *Robert H. Barnes* 341

Ethnicity and Medical Care, edited by Alan Harwood. *Renate Wilson and Max Wilson* 343

Communicating Effectively. The Management Anthology Series, Theme Two: Management Functions, edited by Ann J. Huntsman and Jane L. Binger. *Ruth Lyn Riedel* 344

Health Care Management and Contemporary Perspective, by Seth B. Goldsmith. *John R. Griffith* 345

MEDICAL CARE
March 1982, Vol. XX, No. 3

How Good Are the Data?

Reliability of One Health Care Data Bank

LESLIE L. ROOS, JR., PH.D., NORALOU P. ROOS, PH.D.,
SANDRA M. CAGEORGE, S.M., AND J. PATRICK NICOL, B. COMM.

This study investigates the reliability of the Manitoba Health Services Commission data bank from a variety of perspectives. Emphasizing diagnostic and surgical procedures, the research focuses on those areas in which problems exist and in which the data can be relied upon. Computerized comparisons are stressed, since they can provide cost-effective checks on data quality. One key to performing reliability studies inexpensively is finding information recorded independently: by separate individuals or organizations, at two different times, or in two or more data files. When a particular event has certain logical implications vis-à-vis another, inconsistencies can be located. Face sheet information and data on the performance of major surgical procedures were found to be reliably recorded in the Manitoba data bank. Collapsing ICD-8 diagnoses from medical claims into several categories proved much better than relying upon individual diagnoses. Problems in working with the data included difficulty in distinguishing between closely related surgical procedures and the underreporting of inhospital consultations and nonsurgical procedures.

A MAJOR ISSUE with health care data banks concerns their reliability and validity for a wide spectrum of possible research questions. Other investigators have shown the suitability of Medicare/Medicaid data for the study of ambulatory quality of care,[1] of antibiotic usage[2] and of elective surgery.[3] On the other hand, Demlo et al.[4] have cautioned against the use of information from hospital record abstracts "for detailed research and evaluation."

This article draws its examples from our experience with the Manitoba Health Services Commission data, a comprehensive data bank generally illustrative of systems designed to keep track of patient contacts with providers. This study concentrates on information concerning diagnostic and surgical procedures, identifying those areas in which problems exist and those in which the data can be relied upon.

Following the orientation of Sechrest and Phillips,[5] researchers need to learn the extent to which various sources of bias operate in different kinds of records. As other investigators have mentioned,[6,7] positive results tend to be reported more frequently than negative results. Such an unfortunate bias makes it difficult to build on the experience of others. Knowing what does not work may be as important as knowing what does. Because large data banks seem likely to receive increasing research attention in the future, this discussion of biases in one data bank has wider implications.

Data banks often permit making reliability checks relatively easily. Most re-

---

From the Departments of Business Administration and Social and Preventive Medicine, University of Manitoba, Winnipeg, Manitoba.

This research has been supported by National Health Research and Development Projects Nos. 6607-453-44, 6607-1197-44 and 6607-1156-46, as well as by Research Scholar Award No. 606-1114-48 (to L.L. Roos) and Career Scientist Award No. 607-1001-22 (to N.P. Roos).

Requests for reprints: Leslie L. Roos, Jr., Ph.D., Faculty of Administrative Studies, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

0025-7079/82/0300/0266/$01.05 © J. B. Lippincott Co.

search projects will not be able to afford the large-scale, independent checks performed in the reliability assessment of data collected in the National Hospital Discharge Survey.[1,8] Researchers using data banks need to learn how to perform relatively simple cost-effective checks on data quality. Some of the techniques in this article depend on the availability of a comprehensive data bank, permitting the merging of information collected independently from hospital and physician claims. Such data banks are available in several Canadian provinces.[9,10] They are relatively scarce in the United States, although some investigators have successfully utilized several types of hospital files (Medicare and PAS) for reliability studies.[4] Other techniques suggested here are applicable when just one type of file, either hospital or medical claims, is available.

Comparing Canadian with American information provides a sense of the generalizability of the Manitoba findings to health care data banks elsewhere. Such work both reinforces the applicability of the low-cost techniques described here and suggests ways in which these methods can be expanded.

Data Base

In Manitoba, several separate files—the registration, hospital and medical files—are maintained for payment and control purposes by the Health Services Commission (MHSC). This health insurance data base is characterized by universal coverage (all individuals registered in Manitoba, no matter where care is received). Hospital and medical care are documented in considerable detail because of the fee-for-service payment scheme.

The registration file contains data on the population covered, organized by family registration numbers. With minor exceptions, individuals can be uniquely specified by using birth date, sex and initial identifiers in addition to family registration number.* The medical claims file is the vehicle for paying physicians. Claims are submitted by physicians for services rendered to patients. These claims are filed on a temporal basis and contain the physician's identification number, the patient's identifying information, the patient's residence (checked against the population registry), tariff (determined by whether the visit is a consultation, complete physician examination, regional examination, follow-up visit, and the like), and diagnosis (ICD-8 code). If the visit is a consultation, the number of the referring physician is recorded. As described elsewhere, the claims appear to provide valid data on total patient–physician contact.[11] In a fee-for-service system, claims should not underestimate the volume of physician contact, and, since physician practice profiles are routinely audited with detailed analyses of outlier practice patterns, it is unlikely that a significant amount of overbilling exists.[12]

The hospital file is built on the basis of each admission and contains patient identification, dates of admission and discharge, limited information on services rendered, and one or more diagnoses. Surgical procedures are noted, as well as identification numbers of the surgeon or attending physician, consultants and anesthetist.

The data base generally meets the criteria for a health statistics information system forwarded by White.[13] Although each of these files is maintained separately, with no routine record integration, patient identifiers and physician numbers are unique across all files. Files can be built on individuals (all instances of care received from various physicians and in various hospitals over time), physicians (all care rendered over time), or hospitals (all admissions over time).

---

* The only systemic limitation on identifying individuals concerns two children in a family born in the same year, of the same sex, with the same initials. Claims for such individuals cannot be distinguished until one becomes 19 or marries.

Access to specified portions of the data base has been provided after approval by both the College of Physicians and Surgeons and the Manitoba Health Services Commission. Requests for data preparation are put together by the researchers and Commission staff, meeting to work out problems before the request is formalized. Confidentiality has been maintained through the use of abbreviated identifiers; patient name or address is never specified on the researchers' data. Prior to submission for publication, research papers are reviewed by Commission staff to ensure factual accuracy and to verify that individual patients or physicians cannot in any way be identified from the write-up.

Face Sheet Information

Maximum reliability might be expected when the information being sought is unambiguous and not subject to professional disagreement. As the Institute of Medicine study[8] suggested, "face sheet" data collected as part of a government registry seem likely to be of generally high quality in industralized countries. Information on address and marital status may become outdated, but reported date of birth should remain constant. In Manitoba, date of birth information is obtained for the master registry file and reliably duplicated on the claims file. In a check using survey data from 1971 interviews with over 4,500 elderly,[14] age reported by the interviewee and age calculated from the claims information proved identical in 92 per cent of the cases. Since the former was validated by official documents (birth certificate, citizenship papers, etc.) whenever possible, the reliability is encouragingly high.

Data on major events that are routinely collected in health care institutions also tend to be correctly captured in claims systems.[15] One hundred patients recorded as having died in hospital were identified from hospital claims; searches of vital statistics records corroborated these deaths. Ninety-eight per cent of the deaths (50) in Winnipeg hospitals and all of the deaths (50) occurring in rural hospitals were found in both sources. Such accuracy is dependent on the appropriate reporting system. Thus, pilot studies starting with vital statistics records showed no problems when comparisons were made with deaths reported on hospital claims, but some deaths (both in-hospital and out-of-hospital) were not recorded in timely fashion on MHSC's population registry before 1978. Only when appropriate organizational ties were set up with the Office of Vital Statistics did the reporting of such deaths on the registry become routinized.

Procedures and Diagnoses

Other research problems involve subject matter more ambiguous than births and deaths. Studies of elective surgery must consider the reliability associated with both procedures and diagnoses. The presence or absence of certain operative procedures (tonsillectomy, cholecystectomy, etc.) typically defines operations and control groups in longitudinal outcome studies. Inappropriate selection of diagnosis by the physician may affect the outcome analysis comparing these groups. The situation is complicated, since misclassification may increase or reduce the magnitude of association.[16]

Using several independent sources of information is basic for reliability studies. One type of check looks at procedures and diagnoses across two files, comparing those on the hospital claim with their counterparts on the operating physician's medical claim. Comparing the data on claims submitted by operating physician, anesthetist and assisting physician provides another way to invesigate reliability. In this later case, only one file (the medical claims file) is involved.

Reliability of Procedures

Surgical procedures are reasonably unambiguous events reported on the claims form submitted by the hospital to the Manitoba Health Services Commission for each patient stay. Tables 1 and 2 show general correspondence between hospital claims and surgeons' medical claims. With 722 cases overall, Table 1 contains relatively small numbers in cells suggesting an error. In only 12 cases did the MHSC tariff indicate an abdominal hysterectomy (total) while the ICD-8 operation code noted vaginal hysterectomy (total). The situation was reversed in just eight cases: MHSC tariff indicating vaginal hysterectomy with ICD-8 code noting abdominal. A similar table for gallbladder operations showed 246 cases in which both tariff and ICD-8 code agreed on a cholecystectomy; in 13 cases there was disagreement as to whether or not the common bile duct was explored.

Checks comparing tariffs filed by surgeons, anesthetists and assisting physicians had similarly high levels of agreement. As shown by Table 2, identical tariffs were filed in 91 per cent of the cases involving hysterectomy and in 89 per cent of the cholecystectomy cases. Only a single claim was filed in four and five per cent of the cases, respectively. Thus, direct discrepancies were found in only five (hysterectomy) and seven (cholecystectomy) per cent of the cases. In summary, the MHSC data appear highly reliable in recording the common surgical procedures of cholecystectomy and hysterectomy.

Comparisons of the Manitoba reliabilities with those found in the Institute of Medicine analysis of the National Health Discharge Survey abstracts should take note of the purpose underlying the data banks. As emphasized by a reviewer of this paper, the National Health Discharge Survey seeks to identify a single, principle procedure and/or diagnosis "for tabulation

TABLE.1. 1974 Hysterectomies
ICD-8 Operation Codes on Hospital Claims
and MHSC Tariffs on Surgeon's Medical Claims

ICD-8 Operation Code	4617 Abdominal Hysterectomy (total)	4631 Vaginal Hysterectomy	4621 Hysterectomy (subtotal)	4627 Hysterectomy (radical)	Total	Per Cent
69.1 abdominal (partial)	1	—	5	—	6	1
69.2 abdominal (total)	244	8	2	1	255	35
69.3 abdominal (radical)	—	—	—	4	4	1
69.4 vaginal (total)	12	141	—	—	153	21
67.0–67.9 + 69.2 operations on ovary	288	1	1	2	292	40
68.0–68.9 + 69.2 operations on fallopian tubes	11	1	—	—	12	2
TOTAL	556	151	8	7	722*	100
Per cent	77	21	1	1	100	

* Surgeon's medical claim was missing in 20 cases.

and later derivation of length of stay and discharge rates." The Manitoba data can be used for this purpose, but contain additional information (up to three diagnoses and three operative procedures) to facilitate other administrative and research tasks. The reliability of the Manitoba data may be somewhat higher than that from the Institute of Medicine study, but the methodological differences do not permit a definitive statement. The Institute research reported 85 per cent agreement between National Health Discharge Survey abstract and the reabstract for abdominal surgery, with 76 per cent agreement for gynecological surgery when procedures were coded to two digits. The National Health Discharge Survey's coders had some difficulty in deciding which of several related procedures was to be noted as the principal procedure.†

Reliability of Diagnoses

Several investigators have emphasized that diagnostic reliability approaching 100 per cent cannot be achieved. Koran's research noted substantial disagreement among physicians,[17,18] whether they were eliciting physical signs, interpreting diagnostic tests, making a diagnosis, recommending treatment or evaluating quality of care. Although the diagnosis of some conditions is undoubtedly more reliable than that of others,[19] 60 and 70 per cent levels of agreement are not unusual, even when physicians know that their clinical judgment is being examined.

The reliability of diagnoses appearing on hospital and medical claims can be assessed in several ways. First of all, independently collected information from hospital records can be compared with the MHSC data. Table 3 presents diagnoses among individuals aged 59 years and over who were noted on the hospital records as dying of myocardial infarction in 1974–1976. Diagnoses on the MHSC claims show a high level of agreement with those on the hospital record for four Winnipeg and four rural hospitals. Diagnoses indicating serious coronary problems occur in almost all of the individuals (152 out of 154). In similar fashion, a set of checks using the Manitoba Health Services Commission data and records from one Winnipeg hospital found 37 (95 per cent) of the diagnoses for gallbladder patients to be identical. Levels of agreement using the three-digit ICD codes were higher than those re-

TABLE 2. Comparison of Tariffs on Medical Claims Filed by Surgeon, Anesthetist and Assisting Physician

	Hysterectomy* (N = 742) Per Cent	Gallbladder and Biliary Tract Surgery† (N = 289) Per Cent
All tariffs identical	91	89
Only 1 claim filed	4	5
1 of 2 or more claims filed has discrepant tariff	5	7

* All hysterectomies in MHSC medical claims files for the period April 1–August 31, 1974.

† A 25 per cent sample of gallbladder and biliary tract surgery in MHSC medical claims files for the period April 1–August 31, 1974.

---

† Ignoring ordering discrepancies on hospital discharge claims greatly increases levels of reliability.[4] Removing the assumption that the ordering of the diagnoses contributes to error raises the lowest reliability of any of the 15 Institute of Medicine diagnoses reviewed from 36.8% to 70.6%, with the overall reliability of diagnoses moving from 71.7% to 87.2%. Such high reliability implies that the National Hospital Discharge Survey should prove to be a fruitful research source.

corded for the same conditions in an American study using independently recorded MEDPAR and PAS diagnoses.[3] Such findings suggest that, with appropriate caution, serious operations and life-threatening conditions can be meaningfully investigated using the MHSC data bank.

Time-Sequenced Relationships

Another technique for assessing claims data involves identifying logically time-sequenced relationships. Certain diagnoses should be associated with certain procedures. Thus, tonsillitis diagnoses were found in the year before tonsillectomy, while those of otitis media preceded adenoidectomy.[20] Several such checks were performed in analyzing the risks and benefits of hysterectomy.[21] For example, neither physician visits for pregnancy nor those for cervical disease should logically occur following a total hysterectomy. In fact, we found a 1–2 per cent error rate in the recording of these diagnoses. Other patterns, however, confirmed the general validity of the claims data: for example, menopausal symptoms increased after hysterectomy but not after cholecystectomy.

Intraphysician and Interphysician Agreement

Although recording and logical errors do not appear to present major problems with the hospital data, assessing reliability involves more than checking for simple errors. Psychometricians[22] have traditionally estimated the reliability of measures through the use of the same, or parallel, forms at two closely related time points. As noted in our previous work,[11] this suggests specific questions relevant to the Health Services Commission data:

1. Would the same informed observer (such as a physician) make the same diagnostic judgment twice when observing the phenomenon?
2. Would two different informed observers (such as two physicians) each make the same diagnostic judgment when observing the phenomenon?

These questions were answered using 1974 claims for gallbladder, biliary tract and related problems from patients having cholecystectomies in 1974. All such claims from physician visits taking place within the same 14-day period were used. Sensitivity checks showed only slightly improved results when a shorter (7-day) period between claims was chosen. Table 4 was produced by comparing the first and second claims within the 14-day period. The levels were defined in terms of their implications for gallbladder/biliary tract surgery. Level 1 diagnoses would be taken by many physicians as indicators for surgery. These included ICD-8 574 (chole-

TABLE 3. Comparing MHSC Diagnostic Data with Hospital Records

	Winnipeg (N = 75) Per Cent	Rural (N = 79) Per Cent
Identical MI (myocardial infarction) diagnoses in both MHSC and hospital record (ICD-8 diagnosis of 410)	92	89
MI diagnosis in hospital and related diagnosis* in MHSC record	5	5
MI diagnosis in MHSC and related diagnosis in hospital record	3	4
MI in hospital, unrelated diagnoses in MHSC record	—	3
MI in MHSC, unrelated diagnoses in hospital record	—	—

* Related ICD-8 Diagnoses:

411: Other acute and subacute forms of ischemic heart disease.
412: Chronic ischemic heart disease.
427: Symptomatic heart disease.
413: Angina pectoris.
414: Asymptomatic ischemic heart disease.
428: Other myocardial insufficiency.
429: Ill-defined heart disease.

The sample was made up of individuals aged 59 and over who were discharged with myocardial infarction diagnoses in Winnipeg and rural hospitals in 1974–1976.

TABLE 4. Intraphysician and Interphysician Reliability of Claims for Gallbladder, Biliary Tract and Related Diagnoses

	Intraphysician—Same Physician, 2 Observations Per Cent	Interphysician—Different Physicians, 2 Observations Per Cent
Diagnoses identical or on same level	80	58
Diagnoses one level apart	16	39
Diagnoses two levels apart	4	3
N of pairs	383	319
N of individuals	254	247
N of physicians	182	215, 168

The five cases of pancreatitis (ICD-8 577) were excluded from the analysis.

lithiasis), 575 (cholecystitis, cholangitis), 576 (other disease of gallbladder and biliary tract), 156 (carcinoma of gallbladder and bile ducts—just one case in this sample), and 157 (carcinoma of pancreas—no cases in the sample). Level 2 was made up of the common 785 diagnosis (symptoms referable to abdomen and lower gastrointestinal tract); this diagnosis is widely found both before and after cholecystectomy.[23] Level 3 was a catchall category grouping a number of codes dealing with diseases of the digestive system.

The most reliable single code with a substantial N was ICD-8 574 (cholelithiasis), with 84 per cent intraphysician and 59 per cent interphysician agreement. ICD-8 code 785 and the codes grouped into the "miscellaneous disease of the digestive system" category are responsible for much of the drop in reliability between intraphysician and interphysician comparisons. These two categories were unreliable in themselves, with considerable shifting back and forth between categories. There was some tendency for a less specific diagnosis to be given by the first physician seen, and a more specific judgment (ICD-8 574, 575 or 576) made by the second physician. This is as might be expected in a referral situation. Eliminating claims when one diagnosis was in the miscellaneous disease category raised percentage agreement by 14 per cent for the interphysician category.

These data, in conjunction with respiratory data discussed elsewhere,[11] emphasize the need for grouping three-digit ICD-8 diagnostic information from medical claims. Overreliance on particular individual diagnoses is clearly dangerous, given the instability of the processes underlying clinical judgment. Performing such checks early in the research process is recommended, since the reliability of diagnoses varies according to the organ system and the particular medical problem.[8] These types of checks provide valuable information both on how to group diagnoses and on the extent to which such diagnoses should be used in any specific research project.[‡]

Problems with the Data

Focusing on Certain Procedures

The reliability of some types of diagnoses and operative procedures may be particularly suspect. For example, determining whether or not surgery deprived a woman of ovarian function was relevant for the hysterectomy research. Such a distinction was not possible using surgical tariffs, but ICD codes distinguished between the removal of one ovary and the removal of both ovaries (or of the one remaining ovary). Since ovaries were sometimes removed independently of hysterectomy, data from all women (N = 39) having an ovary taken out in the two years before or after the hysterectomy were checked to see if the coding of ovary removal was logically consistent. In the two cases in which one ovary was taken out following a prior removal, this second operation was coded as the removal of one ovary (ICD-8 67.2) rather than the removal of the remaining ovary (67.4). The diagnostic data did not permit making this critical distinction.

[‡] Under some circumstances, knowing which diagnoses go into which categories may be difficult. Any particular decision may take advantage of chance fluctuations in the data base being examined. In such situations, techniques of cross validation developed in other connections may be useful.[22]

Another surgical study deals with the risks and benefits associated with hip arthroplasty (87.1). A comparison of the procedures coded 87.1 with information written on the admission/separation form filed by the hospital suggested that almost 50 per cent of the procedures coded 87.1 should be excluded from the study (Thompson prosthesis, Mueller plates and pins, etc.). Distinguishing among procedures may be possible using the physician's tariff filed, since payment tariffs are often more specific than operation codes. The problem here is not strictly one of reliability, but rather of not accepting potentially ambiguous categories at face value.

In an effort to assess reliability along the lines of the single-file analysis of surgical procedures, diagnoses on medical claims from surgeons, anesthetists and assisting physicians were compared. Table 6 uses the diagnostic categories outlined in Table 5, highlighting the lack of agreement among diagnoses recorded by physicians involved in hysterectomies. Similar results were found for gallbladder surgery. Twenty-four per cent of the cases showed complete agreement among the different physicians, while 68 per cent concerned level 1 diagnoses pertaining to gallbladder or biliary tract diseases. Anesthetists seldom discriminated among diagnoses within the broad gynecologic and gallbladder-related categories. Most anesthetists merely selected a single diagnosis (such as 574, cholelithiasis), repeating this whenever it might be relevant (i.e., for a cholecystectomy).

Such information suggests that the reliability of a diagnosis might depend on such factors as: 1) the amount of time spent on selecting the diagnosis and entering this information; and 2) the closeness of the physician (both psychologically and technically) to the patient's problem. Given the nature of an anesthetist's work, the physician giving the anesthetic often need not be concerned about the specific diagnosis.

Underreporting

Funding permitted small-scale use of the "reabstracted record" method to check on the reliability of other information collected by the Manitoba Health Services Commission.[24] The reporting of in-hospital consultations and some procedures from the hospital to MHSC is less consistent than the reporting of major operations and

TABLE 5. ICD-8 Diagnostic Categories Used in Gynecologic Research

Gynecologic Diagnoses	
630–677, Y06	Pregnancy
180–184	
091, 098, 612–614, 616, 622	Malignant neoplasm Sepsis
218–221, 615	Benign neoplasm
623, 624, 878	Acquired abnormal anatomy
625	Endometriosis
Y09	Prophylactic sterilization
620, 621, 234	Cervical disease
626	Menstrual disorders
099, 256, 595, 627–629, 752, 786, 939	Other gynecologic disorders

TABLE 6. Comparison of Diagnoses on Medical Claims Filed by Surgeon, Anesthetist and Assistant Physician for Hysterectomy

	Per Cent
All are gynecologic diagnoses in same category*	13
All are gynecologic diagnoses but not in same category	79
Gynecologic and nongynecologic diagnoses	3
All are nongynecologic diagnoses	0
Only one claim filed	4
N = 742	

* Gynecologic diagnosis categories are as described in Table 5; any three-digit code not in that list constitutes a nongynecologic diagnosis.

of diagnoses. The form submitted by the hospital to the Health Services Commission is acceptable when consultations and certain procedures are not indicated. On the other hand, diagnoses must be specified on the form; leaving out surgery would be a major error. Table 7 shows an underreporting of consultations on MHSC claims from both urban and rural hospitals.

Table 8 illustrates a similar underreporting of various procedures on the MHSC records of these cardiac patients. This underreporting is important because relatively few patients undergo such procedures in the first place. Procedures noted on the MHSC record were almost always performed, but the absence of an entry does not mean the procedure was not done. In addition to the procedures shown in Table 8, radiotherapy, shock therapy, hemodialysis, myelography, single blood transfusion, chemotherapy and hyperalimentation were all checked, but these procedures were not frequent enough to document difficulties in the recording. Finally, a check using 39 gallbladder patients from one Winnipeg hospital found a similar underestimation of the amount of cholecystography performed.

Why do reliability problems occur in some cases but not in others? Both technical factors, such as whether an item had to be noted on the MHSC claim, and substantive factors may be involved. Surgical procedures are both recorded in a central position on the hospital claims form and clearly of major medical importance. On the other hand, the miscellaneous procedures are to be checked off toward the bottom of the form and are not required items.

TABLE 7. Comparing MHSC Consultation Data with Hospital Records

	MHSC and Hospital Record			
	Both Yes	Both No	Yes—MHSC No—Hospital Record	No—MHSC Yes—Hospital Record
In-Hospital Consultations—Winnipeg (N = 103 records)				
First consultation	55%	30	—	15
Second consultation	14%	71	3	13
In-Hospital Consultations—Rural (N = 124 records)				
First consultation	12%	82	1	5
Second consultation	2%	98	—	1

Percentages are calculated horizontally.

The sample was made up of individuals aged 59 and over who were discharged with myocardial infarction diagnoses in Winnipeg and rural hospitals in 1974–1976. This table includes data from multiple hospitalizations of the same patient, resulting in a larger N than that of Table 3.

TABLE 8. Comparing MHSC Miscellaneous Procedures Data with Hospital Records

	MHSC and Hospital Record			
	Both Yes	Both No	Yes—MHSC No—Hospital Record	No—MHSC Yes—Hospital Record
Physiotherapy	9%	83	3	4
Inhalation therapy	7%	85	—	8
EKG	98%	—	—	2
Multiple blood transfusion	3%	90	1	6
Cardiac monitoring	69%	10	—	21

Percentages are calculated horizontally; 103 individual records from Winnipeg hospitals were used.

The sample was made up of individuals aged 59 and over who were discharged with myocardial infarction diagnoses in Winnipeg and rural hospitals in 1974–1976. This table includes data from multiple hospitalizations of the same patient, resulting in a larger N than that of Table 3.

Discussion

Several general themes emerge from this review of the quality of the Manitoba data. Finding information recorded independently by separate individuals or organizations, at two different times, or in two or more data files is one key to performing reliability studies inexpensively. Secondly, when a particular event has certain logical implications vis-à-vis another, inconsistencies can be located and their importance assessed.

We have adopted a practical outlook to problems of reliability, emphasizing the sorts of checks that can be done with minimal resources. Although some checks using primary data collection (i.e., going back to the physical hospital records) have been mentioned, computerized approaches have been emphasized. In the process of organizing claims for data analysis, many of these checks can be performed using such packages as SPSS with comparatively little additional effort. The general similarity between the Manitoba data and those from the Institute of Medicine study points up the usefulness of the computer-based techniques described here. Any redundancy—similar information in one or several data files—may permit reliability checks. Thus, Wennberg and Gittlesohn's[3] checks using PAS and Medicare files produced results convergent with the IOM and Manitoba data discussed above—face sheet data and discharge status were seen as the most reliable, procedure data as next in quality, and diagnostic information as presenting the most problems.

Although acceptable reliability of the information contained in large data bases is a prerequisite for meaningful research, researchers need to learn other things about such data banks. Since experimental and quasi-experimental designs require longitudinal information, problems of following individuals through time must be addressed. Tracing individual calls for attention to the reliability and completeness of identifiers. Mobility and life cycle changes complicate the process. Future papers will address this important problem of building individual histories.[14]

Acknowledgments

The authors would like to thank the Manitoba Health Services Commission for facilitating access to claims data. Interpretations and viewpoints contained in this article are the authors' own and do not necessarily represent the official opinion of either the Manitoba Health Services Commission or Health and Welfare, Canada.

References

1. Brook RH, Williams KN. Evaluation of the New Mexico peer review system 1971–1973. Med Care 14(Suppl)1976;1.

2. Federspiel CF, Ray WA, Schaffner W. Medicaid records as a valid data source: the Tennessee experience. Med Care 1976;14:166.

3. Wennberg JE, Gittelsohn AM. A small area approach to the analysis of health system performance. Hyattsville, Md.: Health Resources Administration, 1980. (Final report, HRA contract no. 291-76-0003.)

4. Demlo LK, Campbell PM, Brown SS. Reliability of information abstracted from patients' medical records. Med Care 1978;16:995.

5. Sechrest L, Phillips M. Unobtrusive measures: an overview. In: Sechrest L, ed. New Directions for Methodology of Behavioral Science: Unobtrusive Measurement Today. San Francisco: Jossey-Bass, 1979.

6. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 1976;34:585.

7. Rosenthal R. The "file drawer problem" and tolerance for null results. Psychol Bull 1979;86:638.

8. Institute of Medicine. Reliability of National Hospital Discharge Survey Data. Washington: National Academy of Sciences, 1980.

9. Beck RG, Horne JM. Economic class and risk avoidance: experience under public medical care insurance. J Risk and Ins 1976;43:73.

10. Dyke FJ, Murphy FA, Murphy JK, et al. Effect of surveillance on the number of hysterectomies in the province of Saskatchewan. N Engl J Med 1977;296:1326.

11. Roos LL, Nicol JP, Johnson CF, et al. Using administrative data banks for research and evaluation. Eval Q 1979;3:23.

12. Roos NP. Impact of the organization of practice on quality of care and physician productivity. Med Care 1980;18:347.

13. White KL. Information for health care: an epidemiological perspective. Inquiry 1980;17:296.

14. Roos LL, Nicol JP. Building individual histories with registries: a case study. Winnipeg: Faculty of Administrative Studies, University of Manitoba, 1981.

15. Gittelsohn AM, Senning J. Studies on the reliability of vital and health records: comparison of cause of death and hospital record diagnoses. Am J Public Health 1979;69:680.

16. Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1973.

17. Koran LM. The reliability of clinical methods, data and judgments. N Engl J Med 1975;239:642.

18. Rutkow IM, Gittelsohn AM, Zuidema GD. Surgical decision making: the reliability of clinical judgment. Ann Surg 1979;190:409.

19. Engel L, Davis BJ. Medical diagnosis: present, past, and future. Arch Intern Med 1963;112:512.

20. Roos NP, Henteleff PD, Roos LL. A new audit procedure applied to an old question: is the frequency of t and a justified? Med Care 1977;15:1.

21. Roos NP. Hysterectomies in one Canadian province: a new look at risks and benefits. Winnipeg: Department of Social and Preventive Medicine, Faculty of Medicine, University of Manitoba, 1981.

22. Campbell JP. Psychometric theory. In: Dunnette MD, ed. Handbook of Industrial and Organizational Psychology. Chicago: Rand, McNally, 1976;185–222.

23. Horne JM, Beck RG. Temporal patterns in the use of health services leading to cholecystectomy: a process evaluation using insurance records. Med Care 1978;16:1006.

24. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. In: Sartwell PE, Nathanson N, eds. Epidemiologic Reviews, vol. 2. Baltimore: Johns Hopkins, 1980;210–220.

MEDICAL CARE
March 1982, Vol. XX, No. 3

The Interaction Among the Criteria Physicians Use When Prescribing

ROBERT N. ZELNIO, PH.D.

A randomly selected sample of Iowa physicians was surveyed using a self-administered mail questionnaire that measured the relative importance of prescribing criteria as well as selected physician and practice characteristics. Based on a response rate of 68.3 per cent, it was found that the physicians' criteria for selection of a drug to prescribe could be reduced to two groups. The first group is used to select the generic drug or therapeutic class and is of primary concern to the physicians. Those criteria used in specifying a brand of drug product are of lesser importance to physicians and constitute the second group. It was further discovered that physicians could be segregated into two distinct segments based on the importance of the prescribing criteria. It was not possible, however, to explain the composition of the groups using the physician and practice characteristics measured.

THE DEMAND FOR PHARMACEUTICALS has been described as derived demand.[1] This description aptly conveys the fact that the market for drug products is unique in that the ultimate user or consumer of the product is not the primary decision maker in the purchase process. Rather, this process is controlled by the patient's physician, who acts as an agent in deciding whether or not a purchase will be made and, if so, which product will be purchased. The reality of derived demand in combination with the fear that drug therapy is often inappropriate or employed to excess has prompted many studies of physicians' prescribing habits.

While the suggestion has been made that appropriateness, effectiveness, safety and economy are the four criteria that define rational prescribing,[2] research has revealed that the prescribing process is more complex and often less rational. Prescribing would appear to be a function of five broad factors: 1) the patient's disorder; 2) the product; 3) the manufacturer; 4) the physician; and 5) the physician's environment. Assumedly because its impact is obvious and important, the patient's disorder has not been the subject of studies of the prescribing process per se.

The product characteristics studied center on the drug's effectiveness, risk, cost and pharmaceutical properties. In one form or another, a drug's effectiveness in a given situation has been found to be the most important reason for its being prescribed.[3-5] The risk associated with a drug product is usually studied by means of its potential side effects and has been said to be second to effectiveness in influencing prescribing.[3,4] In a review of relevant research, it was concluded that physicians do not consider costs when prescribing.[6] Nevertheless, the physicians sampled in such studies often report that cost is an important consideration.[4,7] The pharmaceutical properties cited as influencing

---

From the College of Pharmacy, University of Iowa, Iowa City, Iowa.

This research was supported by an NIH Biomedical Research Grant.

Reprint requests: Robert N. Zelnio, Ph.D., Assistant Professor, University of Iowa, College of Pharmacy, Iowa City, IA 52242.

0025-7079/82/0300/0277/$00.95 © J. B. Lippincott Co.

physicians' choice of drugs include the ease with which the drug can be administered and the availability of various dosage forms.[3,5]

The impact of drug manufacturers has been studied from two perspectives. First, it has been shown that the company's image or reputation is associated with physicians' preference among drug products.[8,9] Second, manufacturers have been studied as sources of information about drug products. The information provided by manufacturers through detail men, journal advertising, direct mail advertising, sampling, the Physicians' Desk Reference, at meetings or conventions, and through continuing education materials has been said to be highly influential.[3,7,9-13] While all such sources of drug information appear to be effective in influencing physicians, it should be noted that most studies point to detailing as most effective, followed by journal and direct mail advertising. Furthermore, the impact of these sources seems to diminish as physicians gain experience with drug products.

Education is a physician characteristic that has been the subject of repeated study and has been related to the prescribing process in the form of medical school attended, educational qualifications, year of graduation and extent of postgraduate education.[4,5,10,14,15] The consensus appears to be that more recent graduates with more postgraduate education are better prescribers. Prescribing has also been related to physician age, sex, experience, attitudes, specialty and level of job satisfaction.[4,5,9,10,16,17] The evidence regarding the effect of these characteristics on prescribing is, however, often conflicting, which is partly the result of the fact that some studies relate the characteristics to a measure of prescribing quality while others use the number of prescriptions written.

Noncommercial sources of information and the physician's practice characteristics are two categories of environmental variables that have been examined for their effect on prescribing. In general, noncommercial sources of information, of of which journal articles and colleagues are the most important, tend to be more influential as physicians gain experience with drug products and begin to adopt them on a regular basis.[3,11,18-20] Journal articles are often said to be the second most important source of information and serve to legitimize or reinforce physicians' decisions.[3,4,9,11,12] In addition to increasing over time, the impact of articles appears to be enhanced as physicians read more journals and become more accepted in their communities.[4,12]

The examples and testimonials of peers seem to be the method by which colleagues influence prescribing.[21] Their impact is supported by the finding that the rate of new product adoption is enhanced by the existence of social and professional links between physicians.[18] But while the impact of physicians on their colleagues has been established, other health care colleagues apparently exert little influence on prescribing. For instance, the few studies that have addressed the impact of the pharmacist report that although physicians frequently consult with pharmacists, their influence is weak and may depend on the type of information sought and the size of the city in which physicians practice.[3,7,22] It should be noted, however, that the study of the effect of ancillary personnel on prescribing is sparse and inconclusive, and predates much of the expansion of the traditional pharmacist and nurse roles, not to mention the creation of new paraprofessionals. Yet despite this fact, one study has associated the increased use of ancillary personnel with a measure of better prescribing.[9]

The organization of the physician's practice is one variable that combines the effect of colleagues with other practice characteristics. The opportunity to develop or take advantage of associations with other physicians affects the prescribing process.

Thus it has been found that group practices, hospital affiliations and private versus academic practices influence how a physician prescribes.[5,7,9,23] The physician's patient population is a second practice characteristic that has been found to influence prescribing. While the patients themselves are said to have a weak, although underestimated, impact on prescribing, it has been found that the size of the patient population is inversely associated with the number of prescriptions written and the quality of prescribing, and that physicians having patient populations with high incomes tend to prescribe newer and more expensive drug products.[4,5,9,12,15]

Although most if not all of the studies reviewed have examined aspects of the prescribing process in isolation, many have either concluded or implied that the variables relevant to this process do not always exert their effects independently. For instance, Smith suggests that journal advertisements probably affect prescribing in the presence of other variables.[13] Furthermore, it has been stated that the effects of various sources of information are dependent on the physician's experience with drug products, practice organization, community acceptance and degree of innovativeness.[3,4,7,11,12, 18–20,24]

Despite the observation of these and other such interactions, the studies of the prescribing process have not simultaneously considered the relative importance of the various prescribing criteria, nor have their interdependencies been adequately studied. As a result, the objective of this study was to measure the importance of physicians' prescribing criteria relative to one another and to explore their interdependencies using multivariate analyses.

Methods

This objective was addressed by means of a mail survey of 250 Iowa physicians selected from the *AMA Directory*. The selection process was random with the stipulation that the sample contain only physicians who could be expected to prescribe drugs in the course of their practices. This requirement excluded practitioners with specialties such as radiology, anesthesiology and pathology. Each physician in the sample was sent a questionnaire, cover letter and a self-addressed, stamped return envelope. The survey instrument was designed and administered according to recommendations by Dillman.[25] Nonrespondents received a postcard reminder one week after the initial mailing urging them to complete and return the questionnaire. Replacement questionnaires and return envelopes were sent to the remaining nonrespondents after intervals of three and six weeks.

The importance of eight prescribing criteria was scaled using paired comparisons. These criteria were chosen from among those identified by previous studies. While there may be others or other ways to phrase them, it is believed that the eight criteria used in this investigation accurately represent the most important factors identified to date. The paired comparisons method presents all possible pairs of the criteria in randomized order and requires the respondent to select the most important one from each pair. Thus, each prescribing criterion appeared in seven pairs, for a total of 28 pairs and would have a scale value ranging from 0 to 7. The paired comparison method has the advantages of being easy for the respondent when the items to be rated number less that ten and of providing interval level data.[26] In addition to the paired comparisons, the questionnaire asked each physician for background information on their types of practice, size of practice, estimated patient income and age.

Results and Discussion

Of the 250 questionnaires sent to the physician sample, 23 were returned as undeliverable to the desired respondent for a variety of reasons. A total of 155 completed questionnaires were returned out of the remaining 227, for a total response rate of 68.3 per cent. However, only 118 respondents completed the entire set of paired comparisons, for a usable response rate of 52 per cent. The characteristics of the sample are displayed in Table 1. As can be seen, the sample is primarily composed of middle-aged physicians in private practice with large, middle-income patient populations. With the exception of type of practice, there are no comparable data enabling an examination of the representativeness of the sample depicted by the information of Table 1. In 1978, it was reported that 2,355 physicians in Iowa were office based, while 680 were hospital based. The remainder were in teaching, research, administration, retired or unclassified.[27] Thus, 77.6 per cent of the practicing physicians were office based and 22.4 per cent were hospital based. This compares with 76.2 and 23.9 per cent, respectively, when the clinic physicians of Table 1 are combined with those practicing in hospitals to yield a more comparable institutional classification. While this result provides evidence that the sample is representative of physicians in Iowa on the only variable for which outside information is available, a comparison based on a single variable should not be taken as proof of general representativeness.

The criteria used when evaluating drug products during the prescribing process are listed in Table 2 in order of importance

TABLE 1. Sample Characteristics

Characteristics	Frequency	
	n	Per Cent
Type of practice		
Private, solo, office based	49	36.6
Private, group, office based	53	39.6
Hospital based	20	14.9
Clinic based	12	9.0
Total	134	100.0
Size of practice		
Less than 500 patients	13	10.0
500 to 999 patients	11	8.5
1000 to 1999 patients	34	26.2
2000 to 2999 patients	28	21.5
3000 or more patients	44	33.8
Total	130	100.0
Estimated average annual gross family income of patient population		
Less than $10,000 per year	4	3.2
$10,000 to $19,999 per year	64	50.8
$20,000 to $29,999 per year	53	42.1
$30,000 to $39,999 per year	3	2.4
$40,000 or more per year	2	1.6
Total	126	100.1
Physician's age		
Less than 35 years old	10	7.4
35 to 44 years old	36	26.5
45 to 54 years old	37	27.2
55 to 64 years old	36	26.5
65 or older	17	12.5
Total	136	100.1

TABLE 2. Relative Importance of Criteria Used in the Prescribing Process

Prescribing Criteria	Frequency of Selection	Mean Importance	$t_{(117)}$*
Potential side effects	627	5.3	42.7
Probable efficacy	613	5.2	34.4
Contraindications	611	5.2	32.5
Dosage form availability	268	2.3	16.3
Drug Cost	264	2.2	14.2
Source of drug information	255	2.2	14.0
Frequency of administration	229	1.9	14.0
Manufacturer's reputation	228	1.9	12.1
n		118	

* Significantly different from 0 beyond the 0.05 level.

to the sample physicians along with the frequency and mean number of times each was selected from the paired comparisons. The higher the frequency and mean, the more important each criterion was relative to all others. As stated, the scores on which the means were computed range from 0 to 7, as each criterion appeared in seven pairs. The most important criteria deal with the safety, efficacy and appropriateness of therapy. Of lesser importance are criteria reflecting the mechanics of therapy such as dosage forms, cost and administration. In addition, while previous studies have found the manufacturer and information sources to be highly influential, this study found them to be of little importance. One explanation might be that the manufacturer was separated from the source of information, and since the dominant criteria concerned the safety, efficacy and appropriateness of therapy, perhaps physicians are influenced less by the manufacturer or source of information and more by the content or nature of the information. But it should be noted that each of the means of Table 2 was found to be significantly different from 0 using Student's t-test. This suggests that each of the criteria is of some importance to physicians.

Because the data point to the existence of two distinct groups of prescribing criteria, an F test of the hypothesis that all eight means are equal was performed. As shown in Table 3, this hypothesis can be rejected, lending credibility to the observation that some criteria are more important than others. Also displayed in Table 3 are the results of contrasting a linear combination of the means of the first three criteria with a linear combination of the remaining four. The fact that the difference was significant not only supports the existence of two groups of prescribing criteria but also

TABLE 3. Tests of the Mean Importance of Prescribing Criteria

Source	SS	df	MS	F*
Between groups	2.167.89	7	309.7	119.11
Within groups	2.433.84	936	2.6	
Total	4,601.73	943		

$$t_{(943)} = \frac{1/3(5.2 + 5.2 + 5.3) - 1/5(1.9 + 2.2 + 2.3 + 2.2)}{\sqrt{2.6\,(0.533/118)}}$$

$$= 28.8^*$$

* Significant beyond the 0.05 level.

shows that the first is significantly more important than the second.

The structure of these data was further explored using factor analysis. It was expected that the factor analysis would combine interdependent prescribing criteria along the two dimensions indentified above. But, as can be seen in Table 4, three factors were extracted that have eigenvalues of 1 or more and accounted for all the variance of the data. Although these results do not confirm the two-dimension hypothesis, three limitations of this analysis must be noted. First, there is only one variable that is clearly associated with the third factor. Second, the third factor accounts for a small portion of the total variance. And third, two variables do not load highly on one factor and low on the others as is desired. "Dosage form availability" loads moderately high on two factors, while "probable efficacy" does not load highly on any factor. With larger samples, these limitations might not occur, and a two-factor solution might still be likely. But, even with the limitations, this solution was superior to all others and provides some support for the grouping of criteria observed earlier.

The first factor might be named "brand attributes" because all the variables loading highly deal with individual drug product characteristics, which may vary from manufacturer to manufacturer. In contrast, the variables constituting the second factor relate to the safety, efficacy and appropriateness of therapy in general, and might be used in deciding on the therapeutic class of drug to prescribe. When "probable efficacy" is ignored for purposes of interpretation as is often recommended for variables not loading highly on any factor, the second factor can be named "therapeutic safety." The last factor can easily be named "drug cost." Because first and third factors are composed of variables that are of secondary importance to physicians, while those of factor 2 are most important, it would seem that the physicians in this sample are primarily concerned with the selection of the generic drug or therapeutic class, and only slightly concerned with the

TABLE 4. Factor Analysis of Prescribing Criteria*

Prescribing Criteria	Factors 1	2	3	Commonalities
Frequency of administration	0.70	0.08	0.02	0.49
Manufacturer's reputation	-0.50	-0.00	-0.10	0.26
Source of drug information	-0.54	-0.02	0.11	0.30
Dosage form availability	0.45	0.20	0.38	0.39
Contraindications	-0.19	0.83	-0.09	0.74
Probable efficacy	0.15	0.29	0.00	0.30
Potential side effects	0.06	0.54	0.07	0.30
Drug cost	0.21	0.04	-0.60	0.40
Eigenvalues	4.18	2.50	1.32	
Cummulative variance accounted for	0.52	0.84	1.00	

* After orthogonal rotation.

selection of a particular brand of drug product. Such a decision process would be consistent with the lack of importance of the drug manufacturer.

When the importance ratings of the prescribing criteria were cross-tabulated with the physician and practice characteristics of Table 1, no significant relationships were found. While this can be taken to mean that the importance of the prescribing criteria does not vary with the four variables measured in this study, it would not necessarily mean that the sample is homogeneous with respect to the criteria. As a result, cluster analysis was performed in order to investigate further whether or not the importance of the prescribing criteria varied across physicians. In fact, a two-cluster solution emerged when judged according to the change in the maximum distance within clusters and the ratio of the number of distances within clusters that are less than or equal to the maximum to all distances less than or equal to the maximum.

Table 5 shows the results of cross-tabulating the mean importance of prescribing criteria by cluster. The criteria are listed in order of importance to physicians in cluster 1. As can be seen, 53 physicians make up cluster 1, while 65 are in cluster 2. Two observations can be made immediately. First, the order of criteria within both clusters is different from that of the entire sample, as shown in Table 2. This supports the cluster solution to the extent that it is distinct from the unclustered solution. The second observation is that the listing of the criteria differs from one cluster to the other, with the mean importance ratings differing significantly in five cases. Thus, there are two segments of physicians, and while the drug prescribed is a function of the same criteria for each segment, the importance of those criteria varies from segment to segment. The physicians of cluster 1 are significantly more concerned with the probable efficacy, cost and frequency of administration of therapy, and significantly less concerned with the source of drug information and manufacturer's reputation. Both segments are similar to the extent that the safety, efficacy and appropriateness of therapy continue to be dominant, but the second segment values less objective criteria such as the source of drug information and manufacturer's reputation over more objective criteria related to the mechanics of therapy.

The implications of such clusters are many. For example, pharmaceutical manufacturers should seek to provide the type of information valued by physicians. Both clusters require information on the safety, efficacy and contraindications of therapeutic agents. But the segments differ to the extent that the first would prefer information on the mechanics of therapy, while the second would seem to prefer information on the manufacturer itself and its sources of data. For society, at large, the existence of clusters that imply differing prescribing processes would simply raise questions about the effects of the differing processes. For example, is the quality of therapy higher in one cluster or the the other? If so, strategies should be devised to improve the process of prescribing in the lower quality cluster. One strategy might be for

TABLE 5. Mean Importance of Criteria by Cluster

Prescribing Criteria	Cluster 1	Cluster 2
Probable efficacy	5.7	4.8*
Potential side effects	5.3	5.3
Contraindications	5.2	5.2
Drug cost	3.5	1.2*
Frequency of administration	2.7	1.4*
Dosage form availability	2.5	2.1
Source of drug information	1.5	2.7*
Manufacturer's reputation	1.3	2.5*
n	53	65

* Means are significantly different across clusters at the 0.05 level.

the physicians in such a cluster to delegate responsibility for those criteria that are of less importance to them.

Summary and Conclusions

In an effort to explore the relative importance of physicians' prescribing criteria and the interaction among them, a randomly selected sample of 256 Iowa physicians was surveyed with a self-administered, mail questionnaire. While the questionnaire itself measured physician and practice characterisites and the importance of each criterion relative to all others presented, multivariate analytical tools were used to explore their interdependencies. Although limited by the use of a small, geographically distinct sample, it is believed that this research is unique and contributes to the area of study in at least three ways.

First, no other study has examined the influence of prescribing criteria in a way that allows the investigation of their interdependencies. Previous studies have yielded rankings of criteria based on their individual ratings. Paired comparisons as well as factor and cluster analyses explicitly analyze the importance of each criterion relative to all others.

Second, while each of the criteria was of some importance, the mean importance ratings of the criteria as well as frequency of selection suggest that they might be summarized by two distinct groups. The first group is composed of those criteria that are most important to physicians and might be used to select the generic product or therapeutic class of drug product. The second group includes criteria of lesser importance that would be of use in prescribing a specific brand of drug product. Interestingly, this finding agrees with a report by Caplow and Raymond that physicians claim to select drugs because of their operational characteristics rather than brand, price or popularity.[3]

Third, while it was not possible to differentiate between physicians using the background variables measured in this study, two distinct clusters of physicians do exist based on the relative importance of their prescribing criteria. Most probably, the composition of these segments could be explained by other variables not measured in this research, such as specialty, year of graduation or board certification.

Because of the sample limitations identified, future research should be performed to replicate these findings with larger, more broadly based samples. Such research should be conducted using multivariate methods and should incorporate additional background variables that might be capable of explaining the composition of physician segments. Additionally, the implications of these findings on the prescribing process and marketing of drugs should be examined. Such implications might include the reasons why physicians are relatively unconcerned with brand selection, the possibility for interaction between physicians and pharmacists in the selection of a brand of drug, and whether or not the activities of pharmaceutical manufacturers have or should evolve in light of these findings.

References

1. Smith MC. When friends or patients ask about . . . the cost of drugs. JAMA 1976;235:294.

2. Parish P. Sociology of prescribing. Br Med Bull 1974;30:215.

3. Caplow T, Raymond JJ. Factors influencing the selection of pharmaceutical products. J Mktg 1954;19:20.

4. Lilja J. How physicians choose their drugs. Soc Sci Med 1976;10:364.

5. Miller RR. Prescribing habits of physicians: a review of studies on prescribing of drugs parts IV–VI. Drug Intell Clin Pharm 1973;7:558.

6. Zelnio RN., Gagnon JP. The effects of price information on physician prescribing patterns—a literature review. Drug Intell Clin Pharm 1979;13:158.

7. Lowy DR, Lowy L, Warner RS. A survey of physician's awareness of drug costs. J Med Educ 1972;47:349.

8. Bauer R. Risk handling in drug adoption: the role of company preference. Public Opinion Quarterly 1961;25:546.

9. Becker MH, Stolley PD, Lasagna L, et al. Differential education concerning therapeutics and resultant physician prescribing patterns. J Med Educ 1972;47:121.

10. Stolley PD, Lasagna L. Prescribing patterns of physician. J Chronic Dis 1969;22:397.

11. Miller RR. Prescribing habits of physicians: a review of studies on prescribing of drugs parts VII–VIII. Drug Intell Clin Pharm 1974;8:82.

12. Worthen DB. Prescribing influences: an overview. Br J Med Educ 1973;7:112.

13. Smith MC. Drug product advertising and prescribing: a review of the evidence. Am J Hosp Pharm 1977;34:1221.

14. Garner JD, Smith MC, Mikeal RL. A study of relations of physician specialty, medical school and prescribing patterns. Paper presented to the Academy of Pharmaceutical Sciences, Economic and Administrative Sciences Section, American Pharmaceutical Association, San Francisco, 1971:9.

15. Hemminki E. Review of literature on factors affecting drug prescribing. Soc Sci Med 1975;9:111.

16. Knoben JE, Wertheimer AI. Physician prescribing patterns—therapeutic categories and age considerations. Drug Intell Clin Pharm 1976;10:398.

17. Melville A. Job satisfaction in general practice: implications for prescribing. Soc Sci Med 1980;14A:498.

18. Coleman JS, Katz E, Menzel H. Medical innovation—a diffusion study. Indianapolis: Bobbs-Merrill, 1966.

19. Ferber R, Wales HG. The effectiveness of pharmaceutical advertising: a case study. Journal of Marketing 1958;23:395.

20. Miller RR. Prescribing habits of physicians: a review of studies on prescribing of drugs parts I–III. Drug Intell Clin Pharm 1973;7:496.

21. Cluff LE. The prescribing habits of physicians. Hosp Pract 1967;2:100.

22. Reis L. 81% of high prescribing GP's contact Rx men for new drug data. Am Prof Pharmacist 1968;34:34.

23. Muller C. Outpatient drug prescribing related to clinic utilization in four New York City hospitals. Health Serv Rev 1968;3:142.

24. Winick C. The diffusion of an innovation among physicians in a larger city. Sociometry 1961;24:384.

25. Dillman DA. Mail and telephone surveys: the total design method. New York, Wiley, 1978;187.

26. Hughes GD. Attitude measurement for marketing strategies. Glenview, Ill. Scott Foresman, and Co, 1971;103-04.

27. Wunderdorman LE. Physician Disturbance and Medical Licensure in the U.S., 1978. Monroe, Wis.: Center for Health Services Research and Development, American Association, 1979;108.

MEDICAL CARE
March 1982, Vol. XX, No. 3

Does Cost Information Availability Reduce Physician Test Usage?

A Randomized Clinical Trial With Unexpected Findings

DAVID I. COHEN, M.D., M.SC., PAUL JONES, PH.D.,
BENJAMIN LITTENBERG, B.A., AND DUNCAN NEUHAUSER, PH.D.

Four similar teams of physicians associated with similar inpatient units and randomly assigned patients were used to study the effect of providing physicians with cost information about their use of lab tests and x-rays. Two teams received information about lab test costs, and two teams received x-ray test costs. Test usage fell during the experimental conditions and continued to fall after the experimental period ended in teams in which there was an interested leader.

BECAUSE PHYSICIANS have the power to admit to hospitals, order tests and procedures, prescribe medications, and to influence repeat ambulatory care visits, physicians are believed to control 80 per cent of the costs of medical care. Of the $192 billion spent on medical care in 1978, $76 billion went to hospital care.[1] As much as 25 per cent of the average patient's hospital bill will be for laboratory tests.[2]

We hypothesized that providing physicians with information about the costs of their testing would lead to a reduction in test usage but that this reduction would diminish when the feedback of information ceased. Four internal medicine teams (firms) were provided with feedback over a four-week interval. The sharpest drop in laboratory test usage occurred in the two firms that did *not* receive laboratory test cost data. The results were not consistent with our initial hypotheses, although after the fact they are understandable.

The Firm Structure

Cleveland Metropolitan General Hospital is a 600-bed general and tertiary care hospital, with a major teaching affiliation with Case Western Reserve University, School of Medicine. The general internal medicine service is divided into four separate units called "firms," each consisting of a team of residents and attending physicians responsible for an identical 28-bed inpatient unit, outpatient clinics and similar patient populations based on their random assignment to the four firms. The medical residents stay with their firm throughout their three years, rotating off the firm for subspecialty experience during their second year. Patients are randomly assigned to one of the firms and remain with that firm from then on for their general medical care, inpatient and outpatient. Designed in part to provide an educational focus for general internal medicine and for continuity of care, they are an

---

From the Departments of Medicine, Biometry, and Community Medicine and Epidemiology, Case Western Reserve University, and the Cleveland Metropolitan General Hospital, Cleveland, Ohio.

Reprint requests: David I. Cohen, M.D., M.Sc., Cleveland Metropolitan General Hospital, Room H-331, 3395 Scranton Road, Cleveland, OH 44109.

This study was supported in part by the Cleveland Foundation, Primary Care Funds, the State of Ohio, and the Board of Regents.

0025-7079/82/0300/0286/$00.85 © J. B. Lippincott Co.

extraordinary vehicle for ongoing health services research.[3]

Study Design

We hypothesized that feeding back information about the costs of laboratory and x-ray tests to physicians would lead to reduced use of such tests. When the feedback was discontinued, there would be a decay of the effect and test ordering behavior would rise again, but not as high as its original level.

Two firms were randomly chosen to receive only cost information about lab tests, and the other two firms received cost information only about x-ray tests. Information about actual use of x-rays and lab tests was obtained from all four firms. The x-ray firms served as controls for the lab test firms and vice versa. After much thought, this study design was chosen to allow for the occurrence of a similar Hawthorne effect in all four teams. The hypothesized effects and study design are diagrammatically presented in Table 1, which shows the initial high level of testing (H) during the pre-experimental control period (April 1980). During the experimental period (May 1980), testing would fall to a low level (L) only for the feedback tests (x-rays in firms A and D, lab tests in firms B and C). In the postexperimental follow-up period (June 1–22, 1980), the effect would decay, resulting in a moderate or intermediate level of testing (M).

Literature Review

Rapid increases in expenditures for health care have focused attention on the expanding role of technology as a factor in medical inflation.[4] The diagnostic laboratories and diagnostic radiology are particularly conspicuous consumers of the health care dollar.

Many investigators have found that diagnostic test and x-ray ordering is not optimal.[5–7] Goldberg and Abbott[8] found that even if the physician does order the appropriate test, he or she may not make an appropriate response in terms of additional tests or therapies. Schroeder and O'Leary[9] report that ordering of lab tests and x-rays is higher in university-affiliated hospitals than in community hospitals. In one of the few studies to employ even a rudimentary check on quality of care (by observing mortality rates), Martin et al.[10] found that for coronary care patients in one community hospital, lab usage increased geometrically over 30 years, while mortality in severity-adjusted cases remained stable.

Increasing costs are only part of the burden of diagnostic overkill. Even noninvasive procedures carry health risks for the patient. Schimmel[11] found that almost 3 per cent of the patients in a university medical service suffered ill effects in response to diagnostic tests.

It is not yet clear what factors affect physician ordering patterns. Schroeder et al.[12] report that the quality of the interns on medical service (as ranked by faculty) has no bearing on the quantity of tests ordered. Freeborn[13] was able to correlate lab usage among experienced clinicians to the place

TABLE 1. Hypothesized Results

Team (Firm)		Control Pre-experimental Period (April 1980)	Experimental Period (May 1980)	Follow-up Period (June 1980)
A	L	H	H	H
	Ⓧ	H	L	M
B	Ⓛ	H	L	M
	X	H	H	H
C	Ⓛ	H	L	M
	X	H	H	H
D	L	H	H	H
	Ⓧ	H	L	M

H = High volume of testing, L = Low volume of testing, M = Medium volume of testing; A, B, C, D represent four firms (teams of physicians); L = Volume of lab tests, control conditions; Ⓛ = Volume of lab tests, experimental conditions; X = Volume of x-ray tests, control conditions; Ⓧ = Volume of x-ray tests, experimental conditions.

of their training, age and board certification. He particularly noted that more junior members of the staff tended to follow the example set by the leaders on their service.

Dixon and Laszlo[14] found that a substantial fraction of lab tests ordered had no effect on the course of patient care. In a small sample, they placed a limit on the number of tests allowed each intern. The percentage of tests found to be significant to care jumped from 5 to 23, while total usage fell by 25 per cent. No comment was made on the effect of usage on quality of care.

Many proposals have been put forth to reduce the physician's use of unneeded lab tests and diagnostic x-rays. Many of these center on greater educational effort.[15–17] Griner[18] has had some long-term success with an educational program, but he also used other techniques such as the cost feedback examined in the present study. It has also been reported[16] that the effect of educational programs may decay over the long run.

Many writers have suggested using financial incentives to induce house staff to lower the use of diagnostic facilities. Buck and White[19] report some success in reducing the use of a variety of medical services by more stringent claims review procedures.

A particularly promising approach has been to review physician behavior and feedback information (either privately[20] or publicly[21]) concerning the value of the procedures ordered. Martin et al.[22] found chart review to be superior to financial incentives in eliciting a reduction in test usage by medical residents. It appears that other pressures on the house officer (to be thorough, not to miss a diagnosis, to please the more senior physicians, etc.) are more powerful than financial ones.[5]

It is probably true that most physicians are aware of the importance of keeping hospital costs down. However, it is rare for house officers to know the actual costs involved in even the most simple and routine determinations.[15,23–24,25] Schroeder et al.[26] found that distributing a cost audit to the physician reduced lab test costs by 29.2 per cent. A more recent report[27] showed similar findings, although no attempt was made to correct for the confounding element of length of stay.

Methods

All inpatient laboratory tests, including hematology and chemistry, were recorded for all inpatients. X-ray tests included only chest x-rays, CT scans, ultrasound and liver scans. Charges were based on the hospital price list. Five days a week each firm was given its own lab tests of the preceding day by patient name, type of test ordered and total cost. Monday's report included the weekend's tests. SMA6, 12 and 18 test sets were counted as one test, not as six, 12 and 18 tests; however, their prices differed. These data sheets were handed to the physician.

X-ray tests were compiled weekly and reported by patient name, date, type of test and cost. Each firm's usage data were returned once a week and posted. No comparative data about other firms were posted. At the beginning of the experimental period a brief explanation was given to the senior resident to the effect that test usage was being reported to the firms and test usage was going to be analyzed to see if any changes occurred. During the pre-experimental period test usage was recorded unobtrusively. This recording continued through all three periods of the study. Basic demographic characteristics of the patients were also collected.

Costs and Charges

As in all of the financing of health care, the costs and charges issue is a tangled one. Because much of the cost of lab testing is fixed in the short run, a large reduction in test usage will not be reflected by a large reduction in hospital costs. The staff who draw blood will not be laid off, and the laboratory equipment will remain in place. Long-run changes in testing will very likely result in cost savings to the hospital. However, when lab tests are charged to patients and third parties, then they become clear costs to these payors and immediate savings will occur to them. Test charges have been used here.

Patients in the firms at the start of the control period and remaining in the firms at the end of the follow-up period were excluded. This left 909 patients who entered the four firms as inpatients. Of these patients, 398 overlapped the control-experimental and experimental and follow-up periods, and were excluded. The remaining 511 patients whose stay was completely within one ot the three periods were chosen for primary analysis. Of this group 52 per cent were female, the mean age was 54.5 years, and 66 per cent were white. No significant differences in these characteristics were found between firms or during the three periods of the study, suggesting adequate randomization of patients.*

Findings

Table 2 shows mean values for test usage by the 511 patients in the study. Examination of correlations associated with test usage showed the following results for these 511 patients. The number of first day lab tests is positively correlated with all lab tests ($r = 0.43$, $p < 0.01$), and with total lab charges ($r = 0.43$, $p < 0.01$). Number of lab tests is very strongly related to lab test charges; +0.920 for first day tests and +0.971 for all tests. The length of patient stay is a good predictor of test usage, +0.514 for number of lab tests, +0.483 lab test charges and +0.376 x-ray charges. There is significant but less association with CT use, +0.122 ($p < 0.05$) and ultrasound, + 0.266; length of stay is not correlated with first day lab testing. Our subjective severity-of-illness score was not associated with length of stay, but it was associated with more lab testing (+0.123, $p < 0.05$) for the first day testing, and all tests, +0.157, but not with x-ray use. Higher lab test charges were associated with more x-ray charges (+0.427). Different types of x-ray testing were not correlated with each other. The patient's insurance status had no effect on the use of tests or x-rays.

As predicted, there were no significant differences between firms during the control period. As predicted, the firms that received lab test data reduced their test usage during the experimental period ($p < 0.10$), and it returned to a similar level after the data feedback was removed in the follow-up period.

As predicted, x-ray firms did not change their use of lab tests from control to experimental conditions. The initial volume of testing in the x-ray and lab firms was not significantly different during the control period, but became so during the experi-

TABLE 2. Mean Values and Standard Deviations for Test Usage ($n = 511$)

	Mean	SD
Number of lab tests per admission	20.39	22.79
Lab test charges per admission	$214.81	257.97
X-ray charges per admission	$ 49.80	63.50
Number of lab tests on first day	6.28	5.11
Lab test charges for first day	$ 64.83	60.41

---

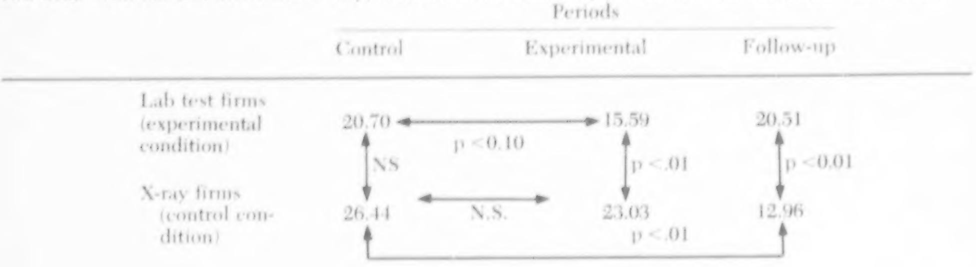
* Patients admitted to the four firms during the study:

There were 302 new patients without overlap between study periods and 209 admissions without overlap for patients who had been seen by the firm physicians prior to this admission, for a total of 511.

There were 398 patients overlapping study periods.

Total number of patients admitted was 909.

FIG. 1. Mean number of lab tests per admission during the control, experimental and follow-up periods for two firms with lab test information only, and two firms with x-ray test information only (total n = 511).


Periods
Control
Experimental
Follow-up
Lab test firms (experimental condition)
X-ray firms (control condition)
20.70
15.59
20.51
p <0.10
NS
p <.01
p <0.01
26.44
N.S.
23.03
12.96
p <.01

mental condition ($p < 0.01$). So far, these findings are consistent with our hypothesis.

What was unexpected was the abrupt and significant drop in the use of lab tests by the x-ray firms in the follow-up period. We had predicted no change, yet it fell to the lowest usage of all—half the volume of the control period. These findings are summarized in Figure 1.

Figure 2 shows the same analysis using laboratory test costs. The differences were not significant, but the pattern is exactly the same as for the number of tests shown in Figure 1. Test usage per patient day showed a similar pattern.

Although the differences in x-ray usage were not significant, there was a decline in x-ray usage during the experimental conditions. This slight decline continued in the x-ray firms but was reversed in the lab test firms. These results are summarized in Table 3.

Unlike the lab test data, which were reported back to the physicians five times a week and usually within 24 hours, the x-ray data were reported back a week later. The lack of immediate feedback to the physicians could explain the smaller effect.

Similar analysis was performed for the smaller group of 302 new arrivals, which showed a similar pattern, and similar results occurred.

Discussion

After the study, physician team (firm) leaders were interviewed concerning their perception of the study. In two firms, which happened to be x-ray firms, firm leaders took an active interest in the study. In one case the full-time chief expressed his interest through informal discussions with the other physicians on the team. In another firm the chief resident required the junior physician with the highest costs to buy the weekly box of donuts for the Friday physician firm conference, a small, symbolic, but visible gesture.

The two passive leaders were in the lab test firms. One of them said that he did not feel it was his role to involve himself in his team's test-ordering behavior. He further said that there is no room for variation in

FIG. 2. Mean laboratory test charges per admission during control, experimental and follow-up periods for two firms with lab test information only and two firms with x-ray test information only (total n = 511).

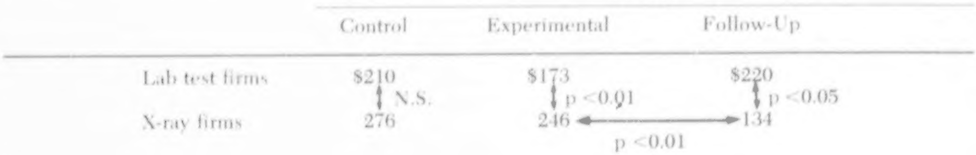

Periods
Control
Experimental
Follow-Up
Lab test firms
X-ray firms
$210
N.S.
276
$173
p <0.01
246
$220
p <0.05
134
p <0.01

TABLE 3. Mean X-ray Charges per Admission during Control, Experimental and Follow-up Conditions for Two Firms with Laboratory Test Information Only and Two Firms with X-ray Test Information Only (Total n = 511)

	Periods		
	Control	Experimental	Follow-up
Lab test firms (control condition)	$55.34	$40.12	$49.35
X-ray firms (experimental condition)	57.37	48.21	45.07

Differences were not significant.

test ordering. A specific type of patient had to have specific tests. Therefore, information on test use and costs could have no effect on physician test-ordering behavior.

Although our findings and other studies show the contrary, his view appears to explain what happened. Cost information must, like all data, be perceived by someone whose mind is prepard to use it, and in the case of a team leader, one who is prepared to convey that message to the team. In the absence of this preparation, information will fail to be useful.

It appears that reductions in laboratory and x-ray tests occurred in the x-ray firms, even though the study design did not provide feedback on laboratory tests and the x-ray feedback was delayed. The greatest change occurred after all feedback had stopped during the follow-up period.

Our findings are quite similar to those of Martin et al.[22] They had three groups of physicians in their study: a control group; a group given a money incentive, $150 if test use dropped by 20 per cent; and a third group with a chart review of 25 per cent of the physicians' patients by a preceptor in weekly discussion groups focusing on tests the preceptor thought to be unnecessary.

We had initially thought of using a money incentive, but our human subjects committee thought it would be unethical to reward physicians not to test. Martin et al.[22] found no change in the control and money reward groups, but a sharp drop in the record review and discussion group, a drop that continued into the follow-up period during which the lowest test usage was found.

Rewards without a "prepared mind" may have led to the lack of change in their money incentive group, as was the case in our lab test groups.

The policy implications seem clear. Simple cost feedback mechanisms will not by themselves assure reduction in test usage. It requires effort to prepare physicians to use these data, an effort that is likely to be cost beneficial because of the large savings that can be achieved by lowering laboratory test charges.

References

1. Gibson RM. National health expenditures 1978. Health Care Finan Rev 1979;1:1.

2. Carels EJ, et al. The physician and cost control. Cambridge, Mass.: Oelgeschlager, Gunn & Hain, 1980;24.

3. Waggoner D, et al. A "firm" system for graduate training in general internal medicine. J Med Educ 1979;54:556.

4. Fineberg HV. Clinical chemistries: the high cost of low-cost tests in medical technology: the culprit behind health care costs? In: Altman SH, Blandon R, eds. Washington, D.C., U.S. Government Printing Office, 1979.

5. Griner PF, Liptzin B. Use of the laboratory in a teaching hospital: implications for patient care, education and costs. Ann Intern Med 1971;75:157.

6. Russe HP. The use and abuse of lab tests. Med Clin North Am 1969;53:223.

7. Abrams HL. The "overutilization" of x-rays. N Engl J Med 1979;300:1213.

8. Goldberg GA, Abbott JA. Explicit criteria for use of lab tests (Letter). Ann Intern Med 1974;81:857.

9. Schroeder SA, O'Leary DS. Differences in lab use and length of stay between university and community hospitals. J Med Educ 1977;52:418.

10. Martin SP, et al. Inputs into coronary care during thirty years: a cost-effectiveness study. Ann Intern Med 1974;81:289.

11. Schimmel EM. The hazards of hospitalization. Ann Intern Med 1964;60:100.

12. Schroeder SA, Schliftman A, Piemme TE. Variation among physicians in use of laboratory tests: relation to quality of care. Med Care 1974;12:709.

13. Freeborn DJ. Determinants of medical care utilization: physicians use of lab services. Am J Public Health 1972;62:846.

14. Dixon RH, Laszlo J. Utilization of clinical chemistry services by medical house staff. Arch Intern Med 1974;134:1064.

15. Hurst JW. Teaching cost-containment to medical house staff. J Med Assoc Ga 1978;67:889.

16. Eisenberg JM. An educational program to modify lab use by house staff. J Med Educ 1977;52:578.

17. Benson ES. Clinical lab utilization and cost of health care. Postgrad Med 1978;63:20.

18. Griner PF. Use of lab tests in a teaching hospital: long-term trends. Ann Intern Med 1979;90:243.

19. Buck CR, White KL. Peer review: impact of a system based on billing claims. N Engl J Med 1974;291:877.

20. Eisenberg JM, et al. Computer based audit to detect and correct overutilization of laboratory tests. Med Care 1977;15:915.

21. Dyck FJ. Effect of surveillance on the number of hysterectomies in Saskatchawan. N Engl J Med 1977;296:1326.

22. Martin AR, et al. A trial of two strategies to modify the test ordering behavior of medical residents. N Engl J Med 1980;303:1330.

23. Wenberg JE. Change in tonsillectomy rates. Pediatrics 1977;59:821.

24. Nagurney JT, et al. Physician awareness of economic factors in clinical decision-making. Med Care 1979;17:727.

25. Robertson WO. Costs of diagnostic tests: estimates by health professionals. Med Care 1980;18:556.

26. Schroeder SA, et al. Use of lab tests and pharmaceuticals: variation among physicians and effect of audit. JAMA 1973;225:969.

27. Henderson D, et al. Hospital cost containment: a little knowledge helps (Abstract). Clin Res 1979;27:279a.

MEDICAL CARE
March 1982, Vol. XX, No. 3

The Effects of Price Information on Physicians' Test-Ordering Behavior

Ordering of Diagnostic Tests

K. MICHAEL CUMMINGS, PH.D., M.P.H., KENNETH B. FRISOF, M.D.,
MICHAEL J. LONG, PH.D., AND GEORGE HRYNKIEWICH, B.A.

This research evaluated the effects of providing physicians with information about the prices of diagnostic tests on their subsequent test-ordering behavior. The study population consisted of 36 second- and third-year residents and 23 clinical faculty in three family practice centers affiliated with the Department of Family Medicine at Wayne State University School of Medicine, Detroit, Michigan. Study participants were asked to review four case studies, each describing a patient with ambiguous symptoms, and to indicate on an attached test order form the tests they would order for each patient. Subjects were randomly assigned either to a group that received test order forms on which the prices of diagnostic tests were printed (price-information group) or to a group that received test order forms with no prices indicated (control group). The study results show that for each of the four cases, the average number of diagnostic tests ordered was significantly lower in the price-information group than in the control group. Our findings also show an average reduction in the cumulative cost of tests ordered per patient of 31.1 per cent related to the provision of price information. The feasibility of regularly providing physicians with price information is discussed and reviewed in light of other approaches that have been developed to modify physician behavior in ordering diagnosite tests.

THE INCREASED USE of hospital and out-of-hospital laboratory testing has contributed substantially to the rise in health care expenditures in the past decade.[1,2] Between 1972 and 1977, expenditures for hospital laboratory tests nearly doubled, from $2.2 billion to over $4 billion.[1] Laboratory testing and radiologic services alone now account for up to 25 per cent of total patient costs at some hospitals.[3,4] The cost of laboratory work both in physicians' offices and in independent laboratories amounted to over $4.2 billion in 1978. Between 1972 and 1977, even though per capita visits to physicians decreased from 5.0 to 4.8, out-of-hospital laboratory tests (excluding x-rays and radiation therapy) increased from 850 tests per 1,000 physician visits to 1,510 tests per 1,000 physician visits.[1]

While much of the increase in the use of diagnostic testing is medically justifiable, there is evidence suggesting that part of the increase in test ordering is unnecessary. In a study carried out at the University of Rochester, Griner and Liptzin[3] examined the relationship between the number of laboratory tests ordered for patients hospitalized with diagnoses of diabetic ketoacidosis and pulmonary edema and

---

From the Department of Cancer Control and Epidemiology, Roswell Park Memorial Institute, Buffalo, New York; the Department of Family Medicine and Community Medicine, Wayne State University, Detroit, Michigan; and the Cost Containment Department, Blue Cross–Blue Shield of Michigan.

This research was supported by NIH grant number 05384-16.

Requests for reprints: K. Michael Cummings, Ph.D., M.P.H., Research Scientist, Department of Cancer Control and Epidemiology, Roswell Park Memorial Institute, 666 Elm Street, Buffalo, NY 14263.

0025-7079/82/0300/0293/$00.95 © J. B. Lippincott Co.

length of stay over a three-year period. Their findings showed a 27 per cent increase in the number of laboratory tests ordered but no corresponding decline in length of stay. Other studies using various criteria to determine the appropriateness of laboratory test ordering have reported substantial over-ordering of such tests as calcium determinations, barium enemas and upper gastrointestinal series.[5-7] In addition, several studies have found that test findings are often ignored by physicians, implying that the tests were not needed to begin with.[3,8,9]

One factor that has been suggested as contributing to the unnecessary ordering of laboratory tests is the physicians' lack of knowledge of the price of tests they order. A number of studies have reported that physicians are generally unaware of the price of commonly ordered diagnostic tests.[10-15] Moreover, these studies have shown a consistent tendency for physicians to underestimate the prices of diagnostic tests.

The origin of this lack of knowledge of prices may very well lie in the fact that health professionals in general are opposed to any consideration of cost in the provision of quality medical care. As Donabedian[16] points out, there is general acceptance of the idea that everything that might contribute positively to the treatment of the patient must be done. However, Donabedian further indicates that while improvements in care will almost certainly result from the increased expenditures, the marginal gain from each new expenditure is likely to diminish. For example, initial studies of cardiac enzymes can yield useful clinical information for diagnosis of myocardial infarction. However, once the diagnosis is made and treatment is initiated, daily cardiac enzyme studies provide little additional useful clinical information.

In this article we report findings from an experimental study designed to evaluate the effects of providing physicians with information about the prices of diagnostic tests on their subsequent test-ordering behavior.

Methods

Study Population

The study population consisted of 36 second- and third-year residents and 23 clinical faculty in three family practice centers affiliated with the Family Medicine Department at Wayne State University School of Medicine in Detroit, Michigan.

Intervention

Study participants were asked to review four case studies, each describing a patient's symptoms and pertinent findings from physical examinations, and to indicate on an attached order form the tests they would order for each patient. A standard test order form was used for all four patients. The cases were constructed to describe common problems seen in adult primary care at a level of ambiguity frequently seen in clinical practice.

Case studies 1 and 2 described patients seen in an ambulatory setting. The first case described a 41-year-old man under marital and job stress who complained of fatigue, weight loss and a variety of mild somatic symptoms in the respiratory, gastrointestinal and urinary tracts. Possible diagnostic considerations included depression, gastritis and alcoholism. The second case described a 56-year-old man with a two-week history of low back pain starting after unusually strenuous activity. The patient's physical examination was most consistent with low back strain, but other causes of the pain were possible.

Case studies 3 and 4 described patients who were admitted to the hospital. The third case described a 57-year-old woman who was an insulin-dependent diabetic with a two-day history of fever, chills and a productive cough. Probable diagnoses included pneumonia and uncontrolled diabetes, among others. The fourth case described a 78-year-old woman with congestive heart failure controlled with digoxin and furosemide who ran out of the diuretic and presented with increasing dyspnea and forgetfulness. Diagnostic considerations included medication noncompliance and acute myocardial infarction.

The four case studies were reviewed by three board-certified internists whose diagnostic impressions matched the intended diagnoses for each case.

Within each family practice center, subjects were randomly assigned either to a group that received test order forms on which the prices of diagnostic tests were printed (price-information group) or to a group that received test order forms with no prices indicated (control group). A stratified random assignment procedure was used to insure a proportionate distribution of residents and faculty in each group. Because prices of diagnostic tests differed in each setting, separate price lists were developed for each family practice center.

Data Collection

One week before data collection began, residents and clinical faculty were sent a letter briefly describing the purpose of the study and urging their participation. Within the next week, subjects received in their office mailboxes the four case studies and test order forms, along with a cover letter describing how to complete the forms. Subjects were instructed to return the completed test order forms to the resident secretary within two weeks. Subjects not returning the materials after two weeks were sent a second set of materials and a letter encouraging their participation. Additional follow-up was carried out by sending daily reminder cards to subjects who had not returned the study materials. Of the 67 subjects contacted, 59 participated in the study (88 per cent response rate).

Dependent Variables

Three variables were analyzed for difference between the price-information group and the control group: 1) a count of the number of diagnostic tests ordered for each case; 2) a count of the number of tests ordered within different categories of tests (11 categories of diagnostic tests were constructed including hematology, urinalysis, chemistry, immunology, radioimmunoassay, microbiology, radiology-plain, radiology-contrast, other radiology, nuclear medicine and other); and 3) the cumulative cost of diagnostic tests selected for each case.

Analysis Plan

Two-sample t-tests were used to evaluate differences in the number of diagnostic tests ordered and the cumulative costs of ordered tests between the price-information group and the control group. Analyses were carried out separately on each case and combined across cases. An F-test was computed to evaluate the homogeneity of variances between the experimental groups. In several instances, the F-test indicated that the variances were unequal. To correct this problem the dependent variables were transformed using either a square root transformation or a logarithmic transformation. The square root transformation was applied to the variables representing the number of tests ordered, and the logarithmic transformation was applied to the variables representing cumulative cost of tests ordered.* After the variables were transformed, there were no statistically significant differences at the $p < 0.05$ level in the variances of the two groups. A nonparametric procedure, the Mann-Whitney U-test, was also used to evaluate differences in the dependent variables between the price-information group and the control group. In no case did the findings from either the nonparametric procedure or from the two-sample t-tests using the transformed variables change the interpretation of the results obtained using the untransformed variables.† The results presented in this article are based on the untransformed data.

---

* When the variates to be transformed include zeros, a technical problem arises, since the logarithmic transformation of zero is negative infinity. To correct this problem, the logarithmic transformation was based on the variable value plus 1.

Results

Figure 1 shows the average number of diagnostic tests ordered for each case by experimental group. For each case, the average number of diagnostic tests ordered is lower in the price-information group than in the control group. Comparison of the number of tests ordered combining all four cases indicates that on the average, physicians in the price-information group ordered 30.8 per cent fewer tests per patient than did physicians in the control group ($p < 0.01$).

Comparison of the average number of tests ordered within 11 different categories of diagnostic test procedures indicates that in all but two instances physicians in the price-information group ordered fewer tests in each category of diagnostic test procedure than did physicians in the control group. For each of the four cases, the largest difference in number of tests ordered is in the category of blood chemistries.

To evaluate whether price information reduces overutilization of unnecessary tests or inadvertently decreases ordering of necessary tests, the price-information group and the control group are compared with regard to the ordering of necessary tests. Three board-certified family practice physicians reviewed each case study to identify the minimum necessary tests that should be ordered for each patient. The necessary tests for case 1 included a complete blood count or hemoglobin, and urine analysis. For case 2, no tests were identified as being absolutely necessary. For case 3, the necessary test identified included a complete blood count with differential, urine analysis, electrolytes, glu-

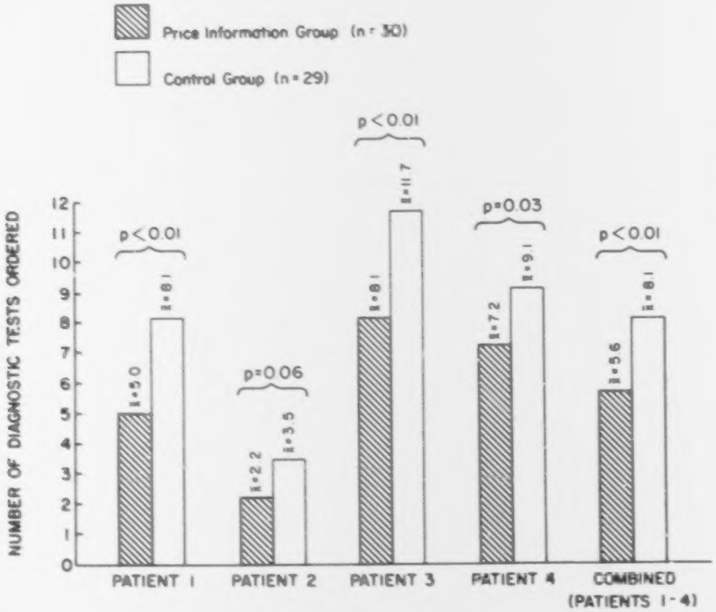


FIG. 1. Average number of diagnostic tests ordered per patient by experimental group.

---

† The results from the two-sample t-tests using the transformed variables and nonparametric tests are available upon request from the author.

cose and chest x-ray. The necessary tests for case 4 included electrolytes, chest x-ray and EKG. Table 1 shows the percentage of physicians ordering the minimum necessary tests for cases 1, 3 and 4 by experimental group. As shown in Table 1, there is little difference in the ordering of necessary tests between physicians in the price-information group and the control group.

The average cost of diagnostic tests ordered for each case by experimental group is shown in Figure 2. For each case, the average cost of tests ordered was lower in the price-information group than in the control group. The difference in the average cost of tests selected was statistically significant at the $p < 0.05$ level for case 1 and case 3, but not for case 2 ($p = 0.09$) and case 4 ($p = 0.09$). Combining the data for all four cases, there is an average reduction in the cumulative cost of tests ordered per patient of 31.1 per cent related to the provision of price information ($p < 0.01$).

The average number of diagnostic tests ordered by clinical faculty and residents is presented in Table 2. As shown in Table 2, in each of the four cases residents ordered more tests than clinical faculty, although none of the differences observed was statistically significant at the $p < 0.05$ level.

The provision of price information affected the test-ordering behavior of residents differently than it did that of clinical faculty. As shown in Figure 3, residents in the price-information group ordered significantly fewer diagnostic tests compared with residents in the control group for case 1, case 2 and case 3, but not for case 4. As indicated in Figure 4, clinical faculty in the price-information group ordered significantly fewer diagnostic tests than did their counterparts in the control group for cases 3 and 4, but not for cases 1 and 2.

Discussion

The findings from this study indicate that providing physicians with information about the prices of diagnostic tests can significantly alter their test-ordering behavior

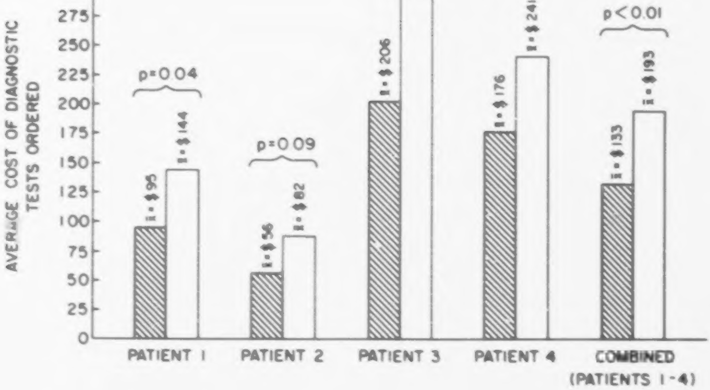


FIG. 2. Average cumulative cost of diagnostic tests ordered per patient by experimental group.

TABLE 1. Percentage of Physicians Ordering Necessary Tests Per Patient by Experimental Group

	Price Information n = 30	Control n = 29
Patient 1	80%	86%
Patient 2	53%	55%
Patient 3	53%	48%

in the direction of ordering fewer tests. Moreover, the provision of price information resulted in a consistent reduction in the number of tests ordered across several different categories of diagnostic tests (although the only consistent, statistically significant reduction in tests ordered was in the category of blood chemistries). Our findings also show a substantial cost savings related to the provision of price information. The cumulative cost of tests ordered per patient was 31.1 per cent lower in the price-information group than in the control group. However, the strong positive correlations found (product-moment correlations ranged from 0.77 to 0.90 for the four cases) between the number of tests ordered and the cumulative cost of tests ordered suggest that most of the cost savings observed was the result of a reduction in the number of tests ordered rather than of selection of less expensive alternative tests.

The differential effects of price information on the test-ordering behavior of residents and clinical faculty are not easily explained. It appears that for residents, the severity or urgency of the patient's condition is an important factor. Price information was more likely to influence the test-ordering behavior of residents for the less acutely ill patients (cases 1 and 2). It is possible that residents are more amenable to the effects of price information when there is less risk associated with not having complete information about the patient's condition. In contrast, for clinical faculty price information had little effect on test-ordering behavior for the commonly seen ambulatory problems (cases 1 and 2). Physicians who have been in practice for some time often have established relatively consistent patterns for diagnosing common conditions and thus are less susceptible to attempts to modify their test-ordering behavior relative to these conditions. A separate explanation is that clinical faculty are more knowledgeable about the prices of commonly ordered diagnostic tests than are residents, and are therefore not as likely to be affected by the provision of price information.

The generalizability of this study's results is limited, both with respect to the sample and by the use of fictitious case studies. The physicians in this study were all family practitioners associated with a

TABLE 2. Average Number of Diagnostic Tests Ordered by Clinical Faculty and Residents

	Clinical Faculty n = 23 x	Residents n = 36 x	Difference	p-value
Patient 1	6.3	6.7	−0.4	p = 0.75
Patient 2	2.5	3.1	−0.6	p = 0.39
Patient 3	9.1	10.4	−1.3	p = 0.21
Patient 4	7.6	8.5	−0.9	p = 0.31
Combined (patients 1–4)	6.3	7.2	−0.9	p = 0.29

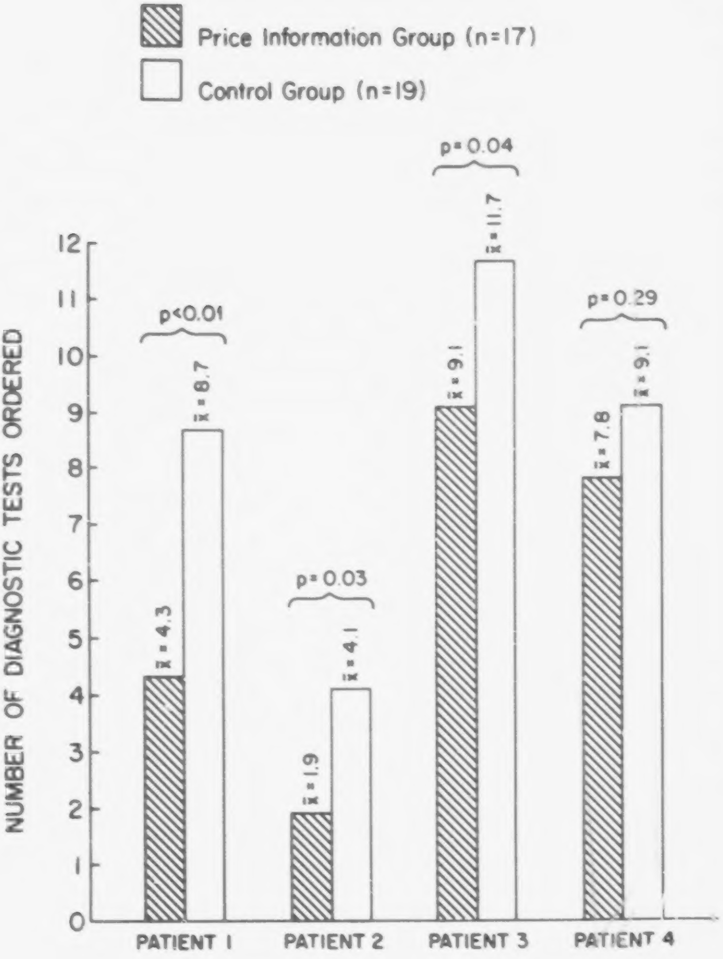


FIG. 3. Average number of diagnostic tests ordered per patient by residents and experimental group.

medical residency training program. It is not clear whether the results obtained with this group of physicians would transfer to other primary care specialty groups or to physicians working in nonacademic settings. It is also not clear whether the results obtained using case study examples accurately reflect physician behavior with their patients. There is some evidence suggesting that the use of written case simulations does provide a good indication of how physicians respond to actual patient problems.[24,25] A further limitation of the case study format used in this research is that it did not allow an opportunity to observe repeat test ordering (i.e., additional test ordering after the results of initial tests are obtained). In a study of the test-ordering behavior of first-year medical residents, Martin et al.[19] reported that repeat testing accounted for about 75 per cent of all laboratory tests ordered for patients in a hospital setting over a one-year period.

Future research should concentrate on evaluating the impact of providing physicians with price information on the number and types of diagnostic tests ordered for actual patient problems and on

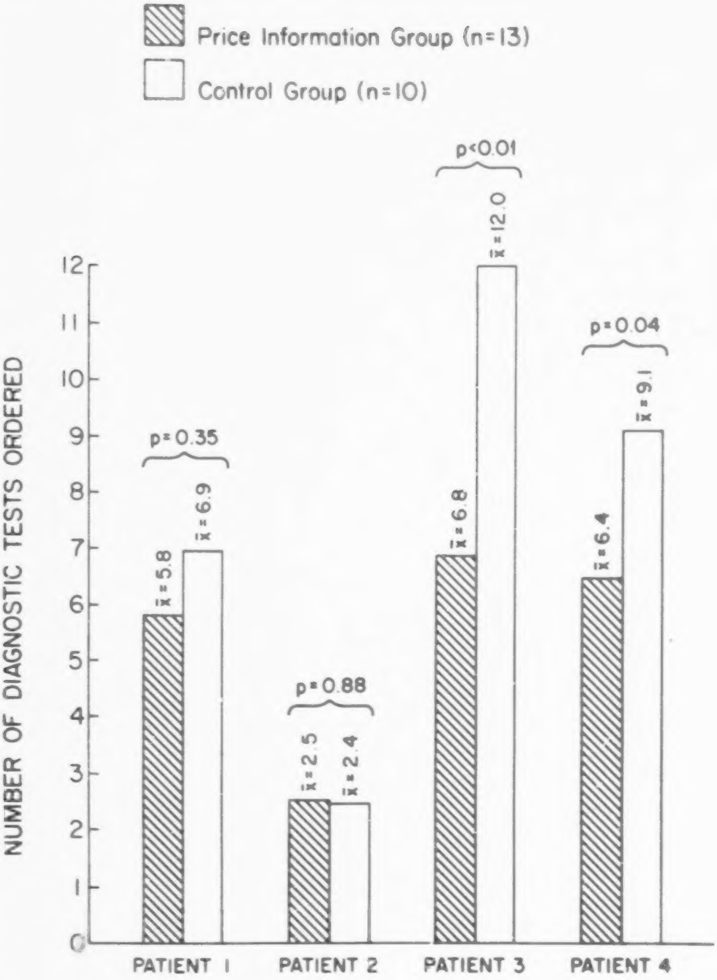


Fig. 4. Average number of diagnostic tests ordered per patient by clinical faculty and experimental group.

the effects of price information in different patient care settings (i.e., hospital, outpatient clinics, private practice, health maintenance organizations). Additional research is also needed to determine which physicians are susceptible to the effects of price information. For example, Schroeder et al.[14] found that a cost audit with feedback to physicians was most effective in reducing the number of tests ordered by "high use" physicians. In our study, the lower variability in the number of tests ordered by physicians in the price information group compared with the control group may reflect a tendency on the part of "high use" physicians to restrain their test ordering when exposed to price information.

Several approaches to modification of physician behavior in ordering diagnostic tests have been reported in the literature. These methods have included setting limits on the number of tests ordered in a day,[8,17] offering incentives to physicians for reduced test ordering,[19] educating physicians about the appropriate ordering of tests,[5,21,22] auditing physician test-ordering practices with feedback to physicians about appropriate test ordering[18–20] and providing daily feedback of patients' bills to the physicians.[14,23] Evaluations of the effectiveness of these different approaches in modifying physicians' test-ordering behavior report reductions in test ordering ranging from 0 to 36 per cent. Overall, auditing of physicians' test-ordering behavior along with feedback concerning appropriate test ordering appears to be the most consistently effective strategy for modifying physicians' test-ordering behavior. Unfortunately, while chart reviews appear to be effective, they are expensive to implement (for example, Martin et al.[19] estimated that it would cost $15,000 per year to conduct six chart review sessions with first-year medical residents on the medical service of a large hospital) and may be resisted if physicians feel that their autonomy and responsibility for patient care are being threatened.

The results of this study suggest that providing physicians with information about the prices of diagnostic tests (before tests are ordered) is an effective strategy for reducing the number of tests that are ordered. While this approach does not deal directly with the problem of the inappropriate ordering of tests, our findings indicate that price information does not affect ordering of necessary tests. In comparison with other approaches for modifying the test-ordering behavior of physicians, providing information on the prices of diagnostic tests is a relatively simple and inexpensive intervention to implement.

References

1. Gibson RM. National health expenditures. Health Care Fin Rev 1979;1:1.

2. Relman AS. The allocation of medical resources by physicians. J Med Educ 1980;55:99.

3. Griner PF, Liptzin B. Use of the laboratory in a teaching hospital: the implications for patient care, education, and hospital costs. Ann Intern Med 1971;75:157.

4. Schroeder SA, O'Leary PS. Differences in laboratory use and length of stay between university and community hospitals. J Med Educ 1977;52:418.

5. Eisenberg JM, Williams SV, Gorner L, et al. Computer-based audit to detect and correct overutilization of laboratory tests. Med Care 1977;15:915.

6. MacEwan DW, Kavanagh S, Chow P, Tishler JM. Manitoba barium enema efficacy study. Radiology 1978;126:39.

7. Morton KI, Sox HC, Wasson J, Duisenberg LE. The clinical value of the upper gastrointestinal tract roentgenogram series. Arch Intern Med 1980;140:191.

8. Dixon RH, Laszlo J. Utilization of clinical chemistry services by medical housestaff. Arch Intern Med 1974;134:1064.

9. Williamson JW, Alexander M, Miller GE. Continuing education and patient care research: physician response to screening test results. JAMA 1967;201:118.

10. Dresnick SJ, Roth WI, Linn BS, et al. The physician's role in the cost containment problem. JAMA 1979;241:1606.

11. Freeman RA. Cost containment. J Med Educ 1976;51:157.

12. Kelly SP. Physician's knowledge of hospital costs. J Fam Prac 1978;6:171.

13. Robertson WO. Costs of diagnostic tests: estimates by health professionals. Med Care 1980;18:556.

14. Schroeder S, Kendeis K, Cooper J, et al. The use of laboratory tests and pharmaceuticals: variation among physicians and effect of cost audit. JAMA 1973;225:969.

15. Skipper JK, Smith G, Mulligan JL, et al. Physicians' knowledge of cost: the case of diagnostic tests. Inquiry 1976;13:194.

16. Donabedian A. Some basic issues in evaluating quality of health care. In: Issues in Evaluation of Health Care. Kansas City, Mo.: American Nurses' Association, 1976.

17. Gray G, Marion R. Utilization of a hematology laboratory in a teaching hospital. Am J Clin Pathol 1973;59:877.

18. Lyle CB, Bianchi RF, Harris JH, Wood ZB. Teaching cost containment to house officers in Charlotte Memorial Hospital. J Med Educ 1979;54:856.

19. Martin AR, Wolf MA, Thibodeau LA, et al. A trial of two strategies to modify the test-ordering behavior of medical residents. N Engl J Med 1980;303:1330.

20. Mitchell JH, Hardacre JM, Wenzel FJ, Lohrenz FN. Cholecystectomy peer review: measurement of four variables. Med Care 1975;13:409.

21. Klein L, Charache P, Johannes R, Lewis C. Effect of physician tutorials on prescribing patterns and drug cost in ambulatory patients. Clin Res 1980;28:296A.

22. Rhyne RH, Gelbach SH. Effects of educational feedback strategy on physician utilization of thyroid function panels. J Fam Prac 1979;8:1003.

23. Henderson D, D'Alessandr E, Westfall B, et al. Hospital cost containment: a little knowledge helps. Clin Res 1979;27:279A.

24. Helfer RE, Slater CE. Measuring the process of solving clinical diagnostic problems. Br J Med Educ 1971;5:48.

25. Williamson JW. Assessing clinical judgment. J Med Educ 1965;40:180.

MEDICAL CARE
March 1982, Vol. XX, No. 3

Physician Sensitivity to the Price of Diagnostic Tests:

A U.S.–Canadian Analysis

JOHN HOEY, M.D., JOHN M. EISENBERG, M.D., M.B.A.,
WALTER O. SPITZER, M.D., M.P.H., AND DUNCAN THOMAS, PH.D.

A questionnaire containing 11 patient management problems was completed by 495 physicians and medical students at an American and a Canadian medical school. Respondents indicated whether they would order a particular diagnostic test in each case, given five different prices for the test. Approximately 25 per cent of attending staff and a higher proportion of residents, interns and clerks responded that they would order the test depending on its price. Approximately 50 per cent of attending staff and smaller proportions of residents, interns and clerks indicated that they would not order the test even if there were no price. Respondents in Montreal were more likely than those in Philadelphia to select a price-sensitive response, the reverse of the expected tendency. Since some tests may be ordered on the basis of price, education of physicians regarding the price of diagnostic tests may alter their use of these services, but a large proportion of tests are ordered because of clinically absolute reasons, which may be insensitive to price.

DURING THE PAST TWO DECADES, the volume and cost of diagnostic investigations have sharply increased. In 1977, it was estimated that clinical laboratories consumed approximately $11 billion in health resources in the United States and that the number of laboratory tests has increased at an annual rate of approximately 15 per cent.[1] Concerns have also been raised regarding the cost and usefulness of radiologic examinations.[2,3] It has been proposed that physician education is an important factor in the use of diagnostic tests, and programs have been initiated to educate physicians about the price of diagnostic studies.[4–6]

The assumption that physicians will use laboratory tests differently as they learn the price of the tests requires a prior assumption that physicians are concerned about the cost of medical care. This implies that

---

From the Department of Epidemiology and Health, McGill University, Montreal, Canada; and the Section of General Medicine, Department of Medicine, and the Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.

This work was supported in part by Health and Welfare Canada (Drs. Hoey and Thomas), the W.K. Kellogg Foundation (Dr. Spitzer) and the National Health Care Management Center at the Leonard Davis Institute of Health Economics, DHEW Grant No HS-02577 (Dr. Eisenberg).

Reprint requests: John Hoey, M.D., Department of Epidemiology and Health, McGill University, 3775 University Street, Montreal, PQ, Canada H3A 2B4.

0025-7079/82/0300/0302/$00.80 © J. B. Lippincott Co.

their use of tests is sensitive to cost. The present study was designed to test whether physicians are sensitive to the price of diagnostic tests and whether physicians differ in this sensitivity depending on their level of training and the health care system in which they practice.

The study was initially performed at McGill University in Montreal, Quebec. Since all hospital and diagnostic test costs in Quebec are fully insured, it was of interest to compare the Quebec findings with those in a fee-for-service setting where costs are born entirely or partly by a least a portion of patients, particularly for outpatient care. Therefore, the study was duplicated at the University of Pennsylvania in Philadelphia, Pennsylvania. Results from McGill and Pennsylvania are presented in this article.

Methods

Study Subjects

The study at McGill was conducted in its five hospital teaching units. At each hospital, random samples of all attending staff, interns and residents in the third residency year were drawn. The sampling fractions were varied to reflect the proportion of physicians in each sample. The sampling fractions also reflected the distribution of physicians among the various specialty departments (Medicine, Surgery, Obstetrics/Gynecology, etc.). Both salaried and nonsalaried faculty were included. Physician administrators, psychiatrists, pathologists, and those on leave of absence or sabbatical were excluded. All family medicine attending staff, interns and residents were included. In addition, a 50 per cent random sample of fourth-year students was drawn. At the Hospital of the University of Pennsylvania, a random sample of students, residents and attending staff was selected in a fashion similar to that used at McGill, except that no family medicine physicians were included.

At McGill, 353 of the 510 subjects returned usable questionnaires (response rate, 69 per cent). At the University of Pennsylvania, 142 of 257 subjects (55 per cent) returned usable questionnaires.

Questionnaire

A questionnaire* was mailed to all study subjects. The questionnaire asked subjects to respond to 11 patient management problems, each of which asked the respondent whether he or she would never order a particular diagnostic test in that clinical situation, would order the test even if it cost more than a specified amount, or would order the test depending on its price. In each case, five different prices of the test were described to determine the physician's perception regarding the relative value of the tests.

The following example is problem 1 from the questionnaire:

> A 32-year-old previously healthy male storekeeper complains of a productive cough of six days, mild fever and malaise. He has rales in the left base, temperature 101°F, pulse 100 regular. He is otherwise well, and the remainder of the physical exam is normal. Would you order a chest x-ray for this patient?
>
> a. Yes, if the chest x-ray cost less than $5.
>
> b. Yes, if the chest x-ray cost between $5 and $15.
>
> c. Yes, if the chest x-ray cost between $15 and 30.
>
> d. Yes, if the chest x-ray cost between $30 and $60.
>
> e. Yes, even if the chest x-ray cost more than $60.
>
> f. No, I would not order a chest x-ray.

---

* Available from the National Auxiliary Publications Service, c/o Microfiche Publications, 304 E. 46th Street, New York, NY 10017.

Table 1 shows how these responses were classified. In the first column we show the responses ("a"–"d") of the subjects who were "price-sensitive." Their responses indicate that they would perform the test only if test price did not exceed a stated amount. Response "e" is shown in the second column and can be termed not cost conscious, or "clinically absolute." These respondents would order the test even if it cost more than the stated amount. Response "f," shown in the third column, is also a "clinically absolute" response. For these respondents the test would not be ordered, even if it entailed no price.

Analysis

A four-way contingency table was used to describe the effects in the data, taking physician groups, cities, questions and response categories as factors. This method produces chi-square tests describing the significance of each combination of effects, considered alone and after adjustment for the other effects that are present in the data. By fitting a log-linear model including only the relevant effects, estimates of the frequency of each response category could be obtained, broken down by each factor separately and adjusted for the others.

Results

With all 11 questions, both cities and all physician groups included for analysis, the price-sensitive options were selected in 33.9 per cent of cases, and the clinically absolute responses, "yes, even if more than the specified amount" and "no, would not order test," in 23.1 per cent and 42.9 per cent, respectively. Though consistent patterns of differences were apparent between questions, cities and professional groups as described below, there remained marked disagreement among individual physicians in response to the same questions in every subgroup. Only one question produced no price-sensitive responses in some subgroups, and even for that question there remained disagreement between "always" and "never" options.

By far the largest effect was a marked variation in the responses to individual patient management problems. For example, overall, 54.5 per cent of respondents to the first question elected a price-sensitive response, and 45.5 per cent a clinically absolute response. In contrast, a patient management problem addressing the use of an electroencephalogram in a 2-year-old girl following a generalized convulsion evoked 11.9 per cent price-sensitive options. These differences among the 11 questions

TABLE 1. Responses for All Patient Management Problems by Professional Category of Respondent

Professional Category	Responses			
	Price Sensitive	Clinically Absolute: "Yes even if more than specified amount"	Clinically Absolute: "No would not order test"	Total
Clerks	37.7%	24.6%	37.7%	904
Interns	37.1	24.1	38.8	1055
Residents	33.5	21.6	44.9	1549
Staff	23.8	27.7	48.5	1763
Total	1788	1220	2263	5271

$\chi^2 = 110.88$; df = 6; $p < 0.001$.

were highly significant ($\chi_{20}^2 = 985.8$, $p < 0.00001$). However, since all physicians in both cities were asked the same questions, these differences could have had no confounding effects on comparisons between cities and physician groups.

On the other hand, differences in response rates produced quite different distributions of physicians in the two cities. In Philadelphia, the proportions of clerks, interns, residents and attending staff were 15.1 per cent, 30.6 per cent, 31.2 per cent and 23.1 per cent, whereas in Montreal they were 18.0 per cent, 15.7 per cent, 28.6 per cent and 37.7 per cent, respectively. As these differences could confound comparisons of response categories, the following results are based on the adjusted proportions obtained from a log-linear model with all significant factors included.

Table 1 shows the adjusted response frequencies in the four physician groups, summed across all patient management problems and across the two cities. These results show two important trends. First, the price-sensitive options were selected most often by the clerks and least often by the staff, with a smooth trend between the interns and residents. Second, among the clinically absolute responses, the "would not order the test" option was selected more commonly by all physician groups, but with increasing frequency among the more senior groups. These overall differences were highly significant ($\chi_6^2 = 110.88$ for partial association, $p < 0.0001$). There were also significant differences between the responses of the physician groups to different questions, but we could find no systematic pattern to such differences.

Table 2 describes the differences in responses between the two cities, summed across problems and physician groups. Montreal physicians chose the price-sensitive option significantly more frequently (35.6 per cent after adjustment) than did Philadelphia physicians (27.7 per cent, $\chi_2^2 = 17.24$, $p < 0.0002$), but the relative frequencies of the "never" clinically absolute options were similar in the two cities. Again there were significant differences between cities in the responses to the different questions, but no consistent pattern of differences could be found.

Discussion

Physicians in many countries are being asked to reduce medical care costs. Laboratory or ancillary investigations are important aspects of such costs. Hence, the mandate for physicians is essentially that they determine whether the laboratory investigation (or other procedure) will yield a benefit to the individual patient that is worth the price.

The results of this study, performed in the United and in Canada, suggest that in

TABLE 2. Responses for All Patient Management Problems in Philadelphia and Montreal

	Responses			
City	Price Sensitive	Clinically Absolute: "Yes even if more than specified amount"	Clinically Absolute: "No would not order test"	Total
Philadelphia	27.7%	31.8%	40.5%	1534
Montreal	35.6	23.3	41.2	3737
Total	1788	1220	2263	5271

$\chi^2 = 17.42$; df = 2; $p < 0.0002$.

both countries a substantial proportion of physicians are willing (at least on a questionnaire) to take price into account and to allow the price of laboratory tests to influence their clinical decisions. This tendency was more marked for clinical clerks and interns, but was also present for attending staff physicians. In 11 patient management problems, an average of almost 25 per cent of attending staff indicated that they would order a laboratory test only if the price for the test were below a specified dollar amount; for clinical clerks and interns, the average so indicating was 37 per cent, with residents occupying an intermediate position. Of note, attending physicians more often indicated that they would not order the test at any price, thus implying that the test would not be clinically useful.

It is possible that attending physicians' reluctance to order tests was due to their clinical experience and to their placing less reliance on laboratory tests than do younger physicians. It has been shown elsewhere that older physicians order fewer laboratory tests and roentgenograms than do younger physicians.[7] It is also possible that younger physicians were most price-sensitive because of recent educational efforts to teach about health care costs. However, it is of note that this study was completed before the initiation of a cost-containment educational program at the University of Pennsylvania. It may also be that a physician's decision to recommend a test as being absolutely necessary incorporated factors other than clinical ones, such as perceived patient demand, habit or defensive medicine. Alternatively, the physician may have been influenced by these considerations more when the test was neither clearly necessary nor unnecessary. In this latter case, the inference is that factors such as defensive medicine would interact with price in determining the physician's decision to order a diagnostic test.

In Quebec, all citizens are completely insured for all investigative procedures. In Philadelphia, few patients have complete insurance coverage for all investigative procedures, particularly for outpatient services. The expected tendency, for Philadelphia respondents to be more price-sensitive, was not found. In fact, Philadelphia respondents were somewhat less likely to indicate that they would order the test depending on its price and more likely to be clinically absolute. It is possible that Philadelphia physicians were more willing to order the test at a price higher than the maximum price listed because they are more accustomed to considering the charge for a service to be its price and since charges for diagnostic services may exceed their true cost. It is also possible that Philadelphia physicians, who may be accustomed to caring for inpatients with extensive third-party coverage, are not aware of the degree to which their patients' diagnostic studies are not covered, particularly when performed in an ambulatory setting. If this were the case, they would be expected to approximate the price-sensitivity of Montreal physicians, but not to fall below it. It is of note that Philadelphia physicians were less price-sensitive than Montreal physicians for diagnostic tests in the seven patient management problems that were characteristically outpatient problems, as well as for those four that were typically inpatient problems.

The results suggest that attending staff, residents, interns and clinical clerks are often willing to take price into account when ordering laboratory tests. Therefore, education of physicians regarding the cost of medical care may influence their use of diagnostic services. However, a large proportion of laboratory tests were considered to be absolutely necessary or unnecessary, regardless of price, particularly by the more senior physicians. Therefore, the potential effect of cost education may be limited. These data must be interpreted with caution since they are based on simulated cases and since all respondents were affiliated with major teaching institutions.

Acknowledgments

The authors wish to thank Dr. A. Ryder-Cook, Ms. S. Kerrouac, Ms. S. Kenmore and Ms. D. Meltzer for their assistance in the field work.

References

1. Fineberg HV. Clinical chemistries: the high cost of low cost diagnostic tests. In: Altman SH, Blendon R, eds. Medical Technology: The Culprit Behind Health Care Costs? Proceedings of the 1977 Sun Valley Forum on National Health: 144–165. (DHEW Publ. No. (PHS) 79-3216.)

2. Abrams HL. The overutilization of x-rays. N Engl J Med 1979;300:1213.

3. Hall FM. Overutilization of radiological examinations. Radiology 1976;120:443.

4. Skipper JK, Smith G, Mulligan JL, et al. Physicians' knowledge of costs: the case of diagnostic test. Inquiry 1976;13:194.

5. Dresnick SJ, Roth WI, Linn BS et al. The physicians' role in the cost containment problem. JAMA 1979;241:1606.

6. Nagurhey JT, Braham RL, Reader GG. Physician awareness of economic factors in clinical decision making. Med Care 1979;17:727.

7. Eisenberg JM, Nicklin D. Use of diagnostic services by physicians in community practice. Med Care 1981;19:297.

MEDICAL CARE
March 1982, Vol. XX, No. 3

Why Veterans Choose Veterans Administration Hospitalization:

A Multivariate Model

WILLIAM FRANK PAGE, PH.D*

Data from a national survey of veterans are analyzed to determine how age, income, health insurance and service-connected disability influence the veteran's choice of Veterans Administration (VA) hospitalization. Log-linear analysis is used to provide a model that expresses in numerically quantifiable terms (odds and odds ratios) the effects of these factors on the choice of VA hospitalization. Moreover, the multivariate nature of the analysis provides the opportunity for assessing the magnitude of these effects simultaneously in the presence of all other effects. The final model shows that age and health insurance act independently of each other and of all other factors. The effect of income is modified by service-connected disability status and vice versa; this interaction of two factors is probably due to VA hospital care eligibility criteria. Overall, the single most important factor affecting a veteran's choice of hospitalization is health insurance; veterans without health insurance have nearly five times greater odds of going to VA hospitals than do those with health insurance—regardless of age, income or service-connected disability status.

ACCORDING TO A National Center for Health Statistics publication,[1] in 1973 the Veterans Administration (VA) accounted for 8.2 per cent of the days of care given in general hospitals other than children's hospitals, and for 11.8 per cent of the days of care in psychiatric hospitals. These figures provide an overall measure of the nationwide magnitude of the VA hospital system.

As the veteran population ages—by 1990 more than half of the U.S. male population over the age of 65 will be veterans[2]—the pressures on the VA hospital system will grow. Furthermore, because veterans over the age of 65 without service-connected disabilities are eligible for VA hospital care by virtue of their age,† the question of whether veterans choose to receive care in VA hospitals may become a critical issue.

In this article we will study the effects of various factors that influence the choice of VA hospitals. We will use a log-linear model analysis because it enables us to express in numerical terms the size of the effects of factors while controlling for the presence of other factors, which is particularly important in analyzing a multifactor issue such as hospital utilization.

---

* Chief, Biometrics Division, Reports and Statistics Service, Veterans Administration Central Office, Washington, D.C.

The opinions of the author do not reflect official Veterans Administration policy and should not be so construed.

Reprint requests: Dr. William F. Page, Chief, Biometrics Division (712), Veterans Administration, 810 Vermont Ave., N.W., Washington, DC 20420.

† The ability to pay for VA hospitalization is not a determining factor for veterans without a service-connected disability if they are aged 65 years or over (see "Choice of Hospital" section for more detail).

0025-7079/82/0300/0308/$01.15 © J. B. Lippincott Co.

Despite the analytic methodology, this article is basically descriptive. We cannot expect to master the complex issue of VA hospitalization in any single analysis or report. Nevertheless, a descriptive study of this sort, providing quantified estimates of the effects of various factors in the presence of other factors, can provide a jumping-off point for future, more definitive studies.

Methodology

Data Source

In 1978, the VA conducted the first National Survey of Veterans.[3] The survey was designed as a follow-up survey to the Bureau of the Census' Current Population Survey (CPS). The CPS is a monthly nationwide survey of about 45,000 households, with the same household included in the sample for several months (4 months in, 8 months out, 4 months in again). Survey of Veterans questionnaires were mailed out to men who reported themselves as veterans in the CPS and who were in the sample for their last time—the so-called "retiring rotation group." Approximately 15,000 questionnaires were mailed out, resulting in a final sample of approximately 11,500 respondents, although this article analyzes only the data from the 1,260 veterans who indicated that they had been hospitalized in the past 12 months. More detail on the survey may be found in the VA report.[3]

Variables

Choice of the variables of interest for the multivariate analysis was made difficult by the small sample size, which made it necessary to limit the number of variables and the complexity of their coding (the more variables and categories, the sparser the data in each category). Hence, the variables in the analysis had to be limited to those with the greatest substantive interest. The group of hospitalized veterans in this analysis was categorized on the basis of five factors (all self-reported): 1) type of hospital (VA or non-VA; persons with multiple hospital episodes were included in the VA category if *any* episode was in a VA hospital); 2) health insurance‡ (yes or no); 3) age (under 45 years, 45–64, 65 and over); 4) income under $10,000, $10,000 and over); and 5) service-connected disability.§

Choice of Hospital

The issue of choice of hospital (VA or non-VA) deserves further discussion since it includes such factors as eligibility and priority for VA care. VA eligibility criteria are relatively complex[4] but may be paraphrased as follows: veterans discharged from active military service under conditions other than dishonorable are eligible; veterans needing hospitalization for service-connected disabilities have top priority; those with service-connected disabilities who need hospitalization for unrelated conditions have second priority; other veterans may be admitted for necessary hospitalization if they state under oath that they are unable to defray the cost of hospital charges elsewhere and if a bed is available; the "inability to pay" requirement does not apply to veterans 65 years of age or older or those in receipt of VA pension.

Given these eligibility and priority criteria, certain hypotheses come to mind. For example, service-connected veterans, since they have higher priority for VA hospitalization, should choose VA hospitals more often. Similarly, non-service-connected veterans under the age of 65 with higher incomes and/or health insurance should choose VA hospitals less often. We cannot say, however, that the latter group is, by definition *ineligible* for VA hospital care, since the inability to defray hospitalization costs is a complex, case-by-case issue that may depend upon such factors as insurance deductibles, coinsurance, limitation of benefits, etc. Hence, we will analyze the data with choice of hospital as the dependent (outcome) variable, without concern for how choice of hospital might depend upon VA hospitalization eligibility and priority.

---

‡ Including Medicare/Medicaid.

§ This factor identifies veterans who have been judged by the VA to have a disability that is service related.

Log-Linear Model Analysis

Log-linear analysis is one of several multivariate methodologies used to analyze data with a discrete dependent (outcome) variable. These methodologies all allow one to express in numerical terms the size of the effects of various factors on choice of hospitalization and quantify those effects in a multivariate manner by assessing the impact of a particular variable, taking into account the presence of all other variables. This last point is particularly important in analyzing a complex multivariate problem such as hospital utilization, which can easily be misinterpreted and misunderstood in a univariate analysis.

There are at least two choices of methodology other than log-linear analysis with the above properties—weighted least squares and logistic multiple regression.[‖] All three methods are "asymptomatically equivalent"[5] (i.e., the differences between results grow smaller as sample sizes grow larger), and the author has chosen log-linear mostly as a personal preference. It should be pointed out, however, that logistic multiple regression will handle continuous and discrete variables, unlike log-linear analysis, which requires all variables to be discrete as does weighted least squares. One can make a substantive case for preferring age as a categorized variable, namely that the age of 65 is a "breakpoint" in defining priority for care, and although there is no substantive argument for income categorization, discrete variables are, in the author's opinion, easier to handle. Analyses with discrete variables offer the advantage of easy computation of odds ratios[8] (definitions and examples are presented in the "Results" section) and, in addition, make use of contingency table frequencies as input, so that one need not process the raw data file more than once, which, although not very important in an analysis of this size, might be a factor in analyzing large files. Availability and ease of use of computer software were also factors in the author's choice of a log-linear methodology.

The analysis was performed using the computer program BMDP3F,[9] with type of hospital treated as a dependent variable. Thus, all models contain a term that fixes the marginal distribution of the independent variables.[10] The actual cell frequencies in the program are the "raw data" in the Appendix and will allow the interested reader to replicate this analysis. [#] The raw data were derived from the weighted survey estimates, although the analysis treated the data as though they had come from a simple random sample.

The last point requires further clarification. Current log-linear methodology is still in developmental stages concerning the analysis of data from other than simple random sampling (the CPS is a multistage sample and so is a complex rather than simple random sample). Because this analysis is primarily descriptive in nature, we have analyzed the data as though they were from a simple random sample; this means that our p-values are only approximate, not a serious shortcoming for this descriptive analysis (a recent article on oc-

---

[‖] When one assumes that the independent variables have a multivariate normal distribution, logistic multiple regression is equivalent to linear discriminant function analysis; this assumption may not be realistic in general and may lead to considerable bias in the linear discriminant function estimates.[6,7]

[#] 0.5 was added to each cell to eliminate zero cells, a traditional practice.

cupational changes, based on data from a supplemental survey to the CPS, takes this approach[11]).

Results

Table 1 shows the percentage of hospitalized veterans with hospital episodes in VA hospitals for the different age, service-connected disability, insurance and income categories. These tables show that older veterans, service-connected veterans, uninsured veterans and poorer veterans are more apt to use VA hospitals. Unfortunately, these simple univariate tables do not provide a complete enough picture. For example, veterans over 65 years of age are more likely to be retired and hence have lower incomes; this increases the VA hospital "market share." On the other hand, these same older veterans are more apt to have insurance (Medicare), which means a reduced VA market share. The only way to disentangle complex analytical situations like this is to perform a multivariate analysis. Log-linear model analysis allows this type of multivariate analysis, which is an important reason for its choice.

To date, no one log-linear model fitting procedure has been adopted as the standard. We have chosen an iterative procedure involving both "stepping-up" (adding variables) and "stepping-down" (deleting variables). The first step is to find a starting model, since looking at all possible models is too costly a process for this data set.

The first step of model fitting was to establish a basic idea of the complexity of the model. Following traditional techniques we fit the model with all one-variable effects¶ (i.e., the effect of a single independent variable on the dependent variable, hospitalization), then all two-variable effects, and all three-variable effects. The results are shown in Table 2.

TABLE 1. Percentage of Hospitalized Veterans with Hospital Episodes in VA Hospitals by Age, Income, Insurance and Service-Connected Disability

Category	Per Cent with VA Hospital Episodes	Total Number of Hospitalized Veterans in Category
Total	16.5	1260.
Age		
Under 45	11.2	373.
45–64	18.6	678.
65 and over	19.2	209.
Income		
Under $10,000	29.3	527.
Over $10,000	7.3	733.
Insurance		
Yes	10.6	1036.
No	43.6	224.
Service-connected disability		
Yes	38.4	211.
No	12.1	1049.

Table 2 shows that the three models in question all fit the data well, with none of the goodness-of-fit chi-squares being significant. In the interest of finding the simplest model, this means that we should start with the "all one-variable effects model."

In the next step, we see if all of the one-variable effects are necessary to the model. We do this by fitting four models, each of which deletes a single one-variable effect. Table 3 shows that each of the independent factors had a significant effect on hospitalization, i.e., each one-variable effect

TABLE 2. First Step in Model Fitting Lack-of-Fit Parameters

Model	χ^2*	df	p-value
All one-variable effects	20.10	18	0.32
All two-variable effects	8.01	9	0.53
All third-variable effects	0.51	2	0.77

* Only likelihood ratio chi-squares are reported in this paper.

¶ All models contained the term that fixed the marginal distribution of the independent variables.

TABLE 3. Second Step in Model Fitting (Step-Down)

Effect Deleted From Model	Change in Chi-Square Due to Deleted Effect		
	X^2	df	p-value
Age	4.88	2	0.09
Service-connected status	46.13	1	<0.0001
Income	53.81	1	<0.0001
Insurance	68.47	1	<0.0001

was significant, although age is significant only at the $p = 0.10$ level.

Since it is clear that the "all one-variable effects" model fits and that no single one-variable effects can be deleted, the next step is to see if any two-variable effects should be added. In Table 4 we look at the effect of adding separately each two-variable effect to the "all one-variable effects" model.

Using the nominal p-value of 0.10, only the income and service-connected status effect is significant. This single two-variable effect is hence added to the model. The next step was to add two-variable effects in addition to income and service-connected status—none was significant. The final step was to double check the model by deleting the effects already in the model. All were significant. This left the final model with all single-variable effects plus the two-variable effect of income and service-connected status.

TABLE 4. Third Step in Model Fitting (Step-Up)

Effect Added to Model	Change in Chi-Square Due to Added Effect		
	X^2	df	p-value
Age and SC status	3.68	2	0.16
Insurance and SC status	0.82	1	0.37
Income and SC status	4.69	1	0.03
Age and insurance	0.78	2	0.68
Income and insurance	0.69	1	0.41
Age and income	2.52	2	0.28

For this model none of the residuals (standardized or Freeman-Tukey deviates) were significant, indicating that the model resembled fairly closely the original data in each cell entry. The fitted data are displayed in the Appendix. The lack-of-fit chi-square for the final model was 15.42, with 17 degrees of freedom ($p = 0.57$), which indicates a well-fitting model.

The final model indicates that all four independent variables had a significant impact upon choice of hospital. Age and health insurance acted independently of each other and of the remaining factors. The service-connected disability factor and the income factor were not independent, so that the effect of the one factor was dependent upon the level of the other. In the usual log-linear terminology, the model included age, insurance, service-connected status, income and the service-connected status by income interaction. We now discuss in turn the effects of each of the factors on the choice of hospital.

Age

Age had a significant effect on the choice of hospital, but it had the weakest effect of the four variables. In general, those veterans under 45 years of age chose VA hospitals less frequently than did the other age groups. Veterans aged 65 years and older tended to choose VA hospitals less often than those between 45 and 64 years of age, but more often than those under 45.

In Figure 1 we present those results in the most straightforward way, using odds and odds ratios. Using the fitted model frequencies, we divided the number of non-VA hospital episodes by the number of VA hospital episodes for any particular age, service-connected status, insurance and income-specific category. This produces the odds that a veteran will be hospitalized in a non-VA rather than in a VA hospital. For example, twice as many non-VA hospi-

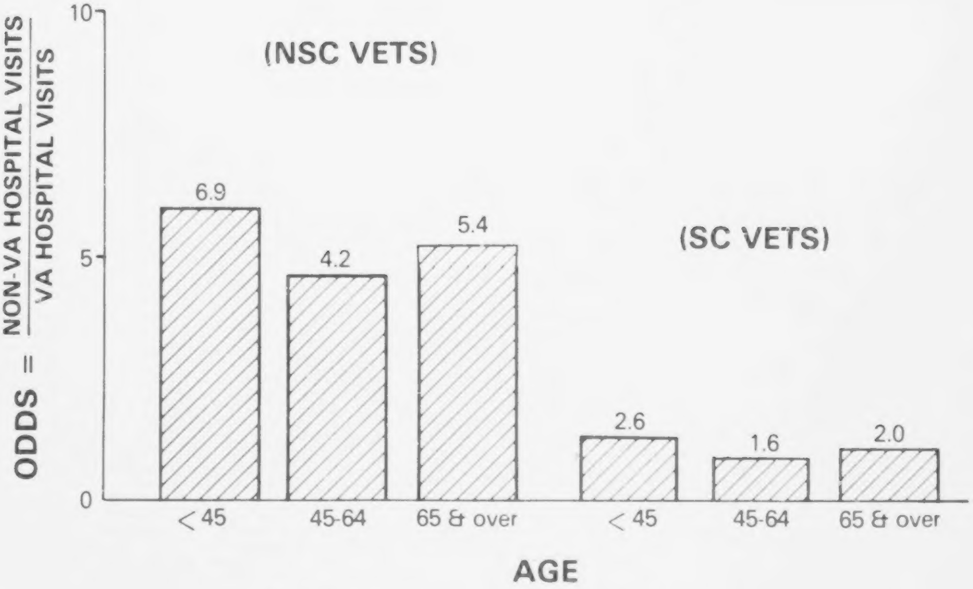


FIG. 1. Odds* of going to non-VA hospital by age.

tal episodes as VA hospital episodes means that the odds are 2 to 1 for non-VA hospitalization. Such odds and even ratios of odds can be expressed mathematically as functions of log-linear model coefficients,[8] but because odds have a simpler substantive interpretation, they should be preferred. Several examples follow.

The following interpretation of Figure 1 leaves out the over-65 age group for simplicity. Reading directly from the graph, it can be seen that service-connected (SC) veterans under the age of 45 are 2.6 times more likely to go to a non-VA than to a VA hospital, and that those 45–64 years of age only 1.6 times as likely. For non-service-connected (NSC) veterans, the odds are 6.9 and 4.2, respectively. Importantly, although these odds differ for service-connected and non-service-connected veterans, the *ratio* of the under 45 years to the 45–64 years odds remains the same for both service-connected and non-service-connected veterans. That is, 2.6 ÷ 1.6 = 1.6 for service-connected veterans, and 6.9 ÷ 4.2 = 1.6 for non-service-connected veterans.

This is an important point. Because the ratios of odds are identical in these two groups, we call it the "age effect," and it is the same regardless of service-connected status. One can also show that this age effect is independent of income and insurance. Thus, we can talk about an age effect without having to specify service-connected status, insurance or income categories, since the odds ratios are the same over these categories.

Health Insurance

In a similar manner we present the data on the health insurance effect. As with age, the effect of health insurance is indepen-

---

* Values shown are for low-income veterans with health insurance; however, the same relationships hold over all income and insurance categories.

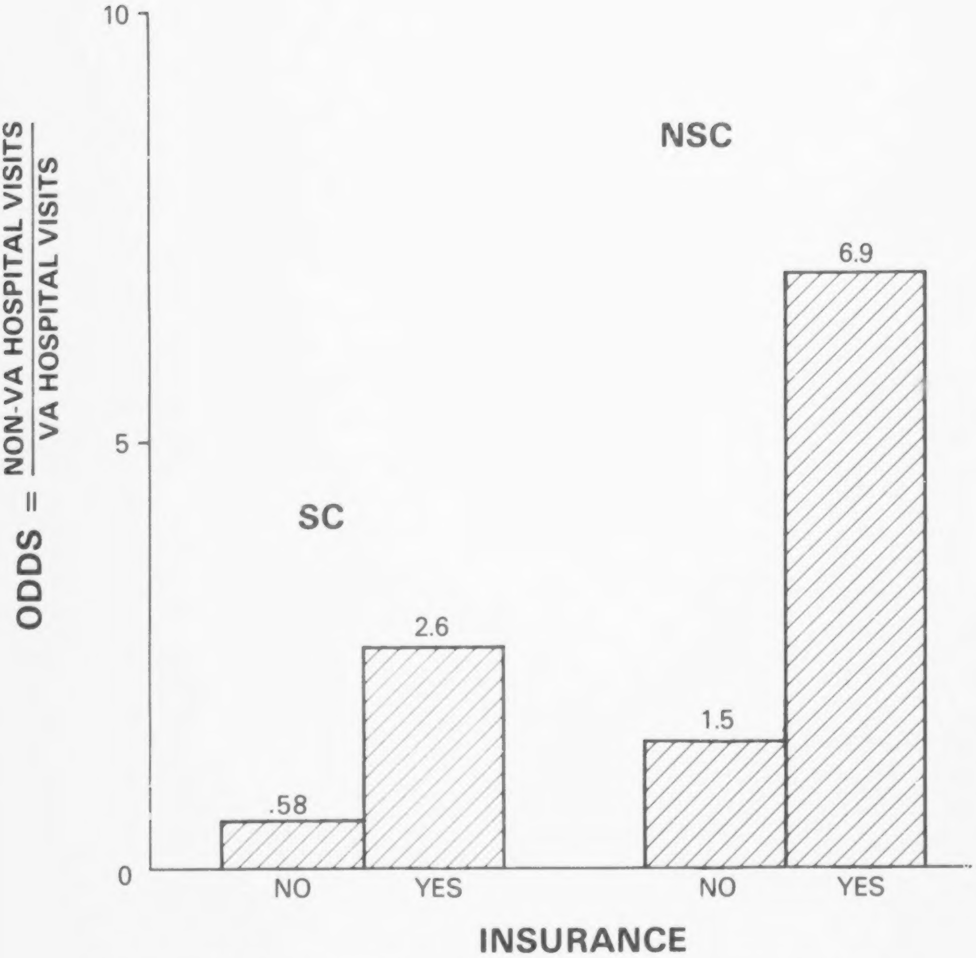


FIG. 2. Odds* of going to non-VA hospital by insurance and service-connected status.

dent of all other factors, but we show service-connected status in Figure 2 for comparability. For service-connected veterans with low income and age under 45 years, the odds of going to a non-VA hospital are 0.58 without insurance and 2.6 with insurance; for non-service-connected veterans, they are 1.5 and 6.9 respectively. Once again we note that the odds ratios are the same for service-connected and non-service-connected veterans, namely, 2.6 ÷ 0.58 = 4.5 for service-connected veterans, and 6.9 ÷ 1.5 = 4.5 for non-service-connected veterans. Hence, we call the common value of the odds ratio (4.5) the "insurance effect."

This odds ratio of 4.5 has an important substantive interpretation. It means that

---

* Values shown are for low-income veterans under age 45 only, however, the same relationships hold over all income and age categories.

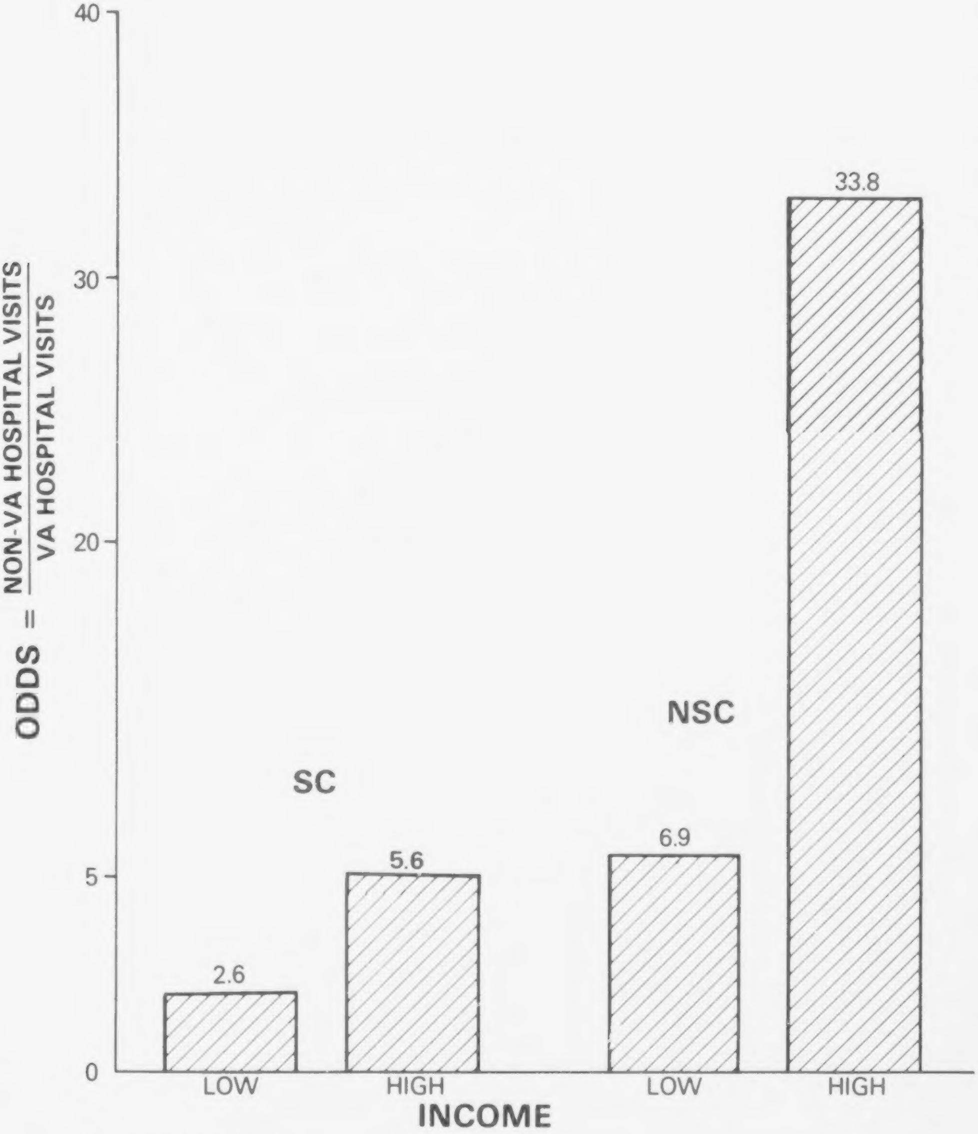


FIG. 3. Odds* of going to non-VA hospital by income and service-connected status.

---

* Values shown are for insured veterans over age 45; however, the same relationships hold over all age and insurance categories.

veterans with health insurance have 4.5 times higher odds of going to a non-VA hospital than do those without health insurance. (Equivalently, those without health insurance have 4.5 times higher odds of going to a VA hospital.) This is a most important effect on the choice of VA hospitals, and it is independent of age, service-connected status and income. That is, in any and every age, service-connected status or income category, those without health insurance show the same 4.5 fold odds ratio favoring VA hospitals. The reader should note that this does not mean that VA hospitals are used more than non-VA hospitals in any absolute sense. It means only that no matter what the odds of non-VA hospital usage are in a particular group with health insurance, they are 4.5 times higher in the corresponding group without health insurance.

Income and Service-Connected Status

The effects of income and service-connected status are not as simple and straightforward to present. The reason is that the effect of income varies with service-connected status and the effect of service-connected status varies with income group. Figure 3 illustrates this graphically.

Figure 3 shows how service-connected status and income act together in the choice of a VA hospital. Among service-connected veterans, high-income veterans have more than double the odds of going to non-VA hospitals (5.6 ÷ 2.6 = 2.2). Among non-service-connected veterans, high-income veterans have five times the odds of going to non-VA hospitals (33.8 ÷ 6.9 = 4.9). In this case the odds ratios for service-connected and non-service-connected veterans are *not* equal as they were for both age and health insurance. This indicates that the income effect is dependent upon service-connected status or, in statistical terms, there is an income by service-connected status interaction. One can also say, equivalently, that the effect of service-connection is dependent upon income. Among low-income veterans, non-service-connected veterans have 2.5 times the odds of going to a non-VA hospital (6.9 ÷ 2.6); among high-income veterans, non-service-connected veterans have six times the odds of going to a non-VA hospital (33.8 ÷ 5.6). A discussion of this interaction is presented in the following section.

Table 5 summarizes our odds-ratio interpretation. Table 5 shows that age and income have less strong effects than do insurance, income and service-connected status. On the whole, health insurance is the strongest single factor. Whereas income has a strong effect among the non-service-connected veterans, it has a less strong effect among the service-connected veterans.

Discussion

Before plunging into a discussion of the fitted model, we should first note some limitations of the data, particularly the fact that the data were self-reported. Self-reporting biases did not appear to be serious in most of the cases for which we could check aggregate survey data against aggregate VA figures (privacy restrictions would not allow us to do any individual record matching). For example, according to VA records, there were 942,000 VA hospital discharges in fiscal year 1977. If we subtract the 45,000 deaths, we calculate 897,000 VA hospital discharges versus the Survey of Veterans estimate of 894,000, a remarkable agreement. Similarly, VA hospital records show that 21 per cent of the discharges (one person may account for multiple discharges) were for persons 65 years of age or older; the Survey of Veterans figure was 18 per cent.

Turning to service-connected disabilities, the overall Survey of Veterans estimate of the number of service-connected veterans agreed well with VA records. However, VA records show that 25 per cent of VA hospital episodes involved service-connected veterans versus the Survey of Veterans figure of 38 per cent. Since the overall Survey of Veterans estimate of service-connected veterans agrees well with VA figures, it must be that either (self-reported) VA hospitalized veterans are more apt to report themselves as having

TABLE 5.

Source	Odds Ratio*	Interpretation
1. Age		
a. Under 45 years vs 45–64 years	1.6	Veterans under 45 years old have nearly twice the odds of using non-VA hospitals as those 45–64 years of age.
b. 65 and over vs 45–64 years	1.3	Veterans 65 and over have about equal odds of using non-VA hospitals as those 45–64 years of age.
2. Insurance Health insurance vs no health insurance	4.5	Veterans with health insurance have nearly 5 times the odds of using non-VA hospitals as those without health insurance.
3. Income a. Service-connected (SC), low income vs high income	2.2	SC veterans with low incomes have more than twice the odds of using VA hospitals as those with high incomes.
b. Non-service-connected (NSC), low income vs high income	4.9	NSC veterans with low incomes have five times the odds of using VA hospitals as those with high incomes.

* The method of model fitting used in this paper does not provide direct estimates of odds ratio reliability; however, all the odds ratios reported are significantly different from 1.0 at the $p < 0.10$ level, so that confidence limits around them exclude the value 1.

a service-connected disability or that (self-reported) service-connected veterans are more apt to remember and report VA hospital episodes. Whatever the case, this overreporting would tend to exaggerate the VA hospital market share among service-connected veterans.

Turning our attention to the fitted model, perhaps the most striking aspect of the log-linear model analysis is the simplicity of the model. With the exception of service-connected disability status and income, the other factors acted independently. Hence, age and health insurance had independent effects regardless of service-connected status and income.

The service-connected status by income interaction is not difficult to explain. Service-connected veterans are given the highest priority for treatment in VA hospitals, while non-service-connected veterans may enter a VA hospital only if they state under oath that they are unable to defray the cost of hospital charges elsewhere (the "inability to pay" requirement does not apply to veterans 65 years of age and older or those in receipt of VA pension).

Thus, it is easy to see why service-connected status and income effects interact. Both high- and low-income service-connected veterans have high priority for VA hospital care, while only low-income non-service-connected veterans have priority for VA hosptial care. Thus, for non-service-connected veterans we saw that the odds that they will go to a non-VA hospital are much higher for high-income than for low-income non-service-connected veterans simply because of "entitlement." It is a bit puzzling that no age by service-connected status interaction was found, since non-service-connected veterans over 65 years of age are not required to meet the "inability to pay" requirement as explained above. This failure to detect the service-connected status by age interaction may be due to the small sizes in the 65 years and over age group.**

Conclusion

The findings of this paper may provide useful though limited information to policy makers. Although identification and interpretation of policy-relevant findings are jobs best left to others more expert, the author has been asked by the reviewers at least to speculate on some areas of policy interest. Let us take, as an example, National Health Insurance (NHI).

According to the model, which ignores all real-life factors save age, income, insurance and service-connected status, a national health plan that acts like third-party insurance or Medicare/Medicaid could seriously erode the VA hospital market share. This is because veterans with existing insurance (our definition included Medicare/Medicaid) are far less likely to choose VA hospitalization, regardless of age, income and service-connected status. This speculative hypothesis is partly confirmed by early Medicare experience, which showed a drop-off in VA hospital usage rates in veterans aged 65 years and over just after Medicare went into effect. This original drop-off was only temporary; VA hospital usage rates among older veterans are now back to pre-Medicare levels. Note that, although the model predicts a sharp drop in VA hospital usage rates in the NHI scenario above, the nondynamic model cannot address the question of a subsequent return to previous usage rates.

On the other hand, one can imagine a national plan that would cover veterans by simply eliminating the distinctions in admission priority between veterans, especially between service-connected and non-service-connected veterans. Such a plan would, according to the model, increase the VA hospital market share, especially among high-income veterans, and would have the opposite effect of the first scenario.

In all of this speculation, however, one must remember that the results presented have come from a one-time static model. Without a dynamic component to the model, predicting future events is somewhat adventuresome. Recall in particular that the model predicts a usage rate drop in the first scenario but can say nothing about subsequent usage patterns. Nevertheless, the model does help us to understand the current situation. Others have found that veterans who use VA hospitals are poorer, less likely to have health insurance and more apt to have a service-connected disability.[12] What we have added is a documentation of the relative strengths of these factors and their interrelationships.

References

1. National Center for Health Statistics. The nation's use of health resources 1979. Washington, D.C.: USDHEW, Public Health Service, 1980. (DHEW publication in. 80-1240.)

2. Committee on Veterans' Affairs, United States Senate. A report on the aging veteran. Washington, D.C.: U.S. Government Printing Office, 1978. (95th Congress, 2d Session, Senate Committee Print No. 12.)

3. Reports and Statistics Service, Office of Controller, Veterans Administration. National Survey of Veterans. Washington, D.C.: U.S. Government Printing Office, 1980.

4. Veterans Administration. Federal Benefits for Veterans and Dependents. Washington, D.C.: U.S. Government Printing Office, 1980.

5. Rao CR. Linear Statistical Inference and its Applications. 2nd edition. New York: Wiley & Sons, 1973:417.

6. Halperin M, Blackwelder WC, Verter JI. Estimation of the multivariate logistic risk function: A comparison of the discriminant function and maximum likelihood approaches. J Chron Dis 1971;125:158.

7. Press SJ, Wilson S. Choosing between logistic regression and discriminant analysis. J Am Stat Assoc 1978;699.

8. Page WF. Interpretation of Goodman's loglinear model effects, an odds ratio approach. Soc Meth Res 1977;5:419.

9. Dixon WJ, Brown MB, ed. BMDP-77 Biomedical Computer Programs P Series. Berkeley: University of California Press, 1977.

** A preliminary analysis of the next year of data shows the presence of this interaction.

10. Fienberg SE. The Analysis of Cross-Classified Categorical Data. Cambridge, Mass.: MIT Press, 1977; chapter 6.

11. Mare RD. Social background and school continuation decisions. J Am Stat Assoc 1980;75:295.

12. Committee on Veterans' Affairs, United States Senate. Study of health care for American veterans. A report prepared by the National Academy of Sciences, National Research Council. Washington, D.C.: U.S. Government Printing Office, 1977. (95th Congress, 1st Session, Senate Committee Print No. 4.)

Appendix

TABLE 1A. Fitted Data*

Income	Health Insurance	Age Group	Service-connected	Hospitalized	
				VA	Non-VA
Low	Yes	Under 45	Yes	2.55	6.57
Low	Yes	Under 45	No	8.57	58.96
Low	Yes	45–64	Yes	15.15	23.98
Low	Yes	45–64	No	23.41	99.04
Low	Yes	65 and over	Yes	7.43	14.97
Low	Yes	65 and over	No	18.69	100.73
Low	No	Under 45	Yes	2.27	1.31
Low	No	Under 45	No	19.77	30.42
Low	No	45–64	Yes	28.66	10.15
Low	No	45–64	No	20.16	19.08
Low	No	65 and over	Yes	3.44	1.55
Low	No	65 and over	No	10.41	12.56
High	Yes	Under 45	Yes	3.48	19.35
High	Yes	Under 45	No	5.96	201.51
High	Yes	45–64	Yes	12.37	42.30
High	Yes	45–64	No	15.60	324.26
High	Yes	65 and over	Yes	1.74	7.58
High	Yes	65 and over	No	1.24	32.87
High	No	Under 45	Yes	0.93	1.16
High	No	Under 45	No	2.21	16.72
High	No	45–64	Yes	8.15	6.24
High	No	45–64	No	6.54	30.41
High	No	65 and over	Yes	1.04	1.01
High	No	65 and over	No	0.31	1.86

* Cell values divided by approximate average weight.

TABLE 2A. Raw Data*

Income	Health Insurance	Age Group	Service-connected	Hospitalized VA	Hospitalized Non-VA
Low	Yes	Under 45	Yes	0.85	7.27
Low	Yes	Under 45	No	10.07	56.47
Low	Yes	45–64	Yes	16.91	21.21
Low	Yes	45–64	No	23.75	97.70
Low	Yes	65 and over	Yes	3.66	17.74
Low	Yes	65 and over	No	19.74	98.68
Low	No	Under 45	Yes	1.76	0.82
Low	No	Under 45	No	16.04	33.14
Low	No	45–64	Yes	31.84	5.98
Low	No	45–64	No	17.25	20.99
Low	No	65 and over	Yes	1.48	2.52
Low	No	65 and over	No	11.16	10.82
High	Yes	Under 45	Yes	5.23	16.60
High	Yes	Under 45	No	4.90	201.59
High	Yes	45–64	Yes	7.85	45.82
High	Yes	45–64	No	14.03	324.83
High	Yes	65 and over	Yes	2.43	5.89
High	Yes	65 and over	No	0.77	32.33
High	No	Under 45	Yes	0.64	0.45
High	No	Under 45	No	2.27	15.66
High	No	45–64	Yes	7.51	5.88
High	No	45–64	No	6.89	29.05
High	No	65 and over	Yes	1.06	0
High	No	65 and over	No	0	1.17

* Cell values divided by approximate average weight.

MEDICAL CARE
March 1982, Vol. XX, No. 3

Satisfaction With Maternity Care:

A Matter of Communication and Choice

DEBORAH A. SULLIVAN, PH.D.,* AND RUTH BEEMAN, R.N., C.N.M., M.P.H.†

The purpose of this study was to determine the level of satisfaction with maternity care and whether satisfaction was related to the patterns of communication between caretakers and patient and to specific clinical procedures used during labor and delivery. Data, gathered in a statewide survey, indicated widespread satisfaction with prenatal, labor and delivery care. Level of satisfaction was related to both the quality of communication and the fulfillment of preferences about the management of labor and delivery. There was a general consensus among respondents about the desirability of childbirth coaches, controlled breathing and relaxation techniques, and freedom to move around and the undesirability of medication rendering one unconscious for delivery. There was considerable variation in desire concerning medication during labor, use of a fetal monitor, presence of family members and choice of atmosphere. The implications for alternatives in maternity care are discussed.

JUDGING BY THE FLOOD of articles in the popular women's press, the last five years have been a time of growing unrest with standard maternity care. Numerous authors have questioned such practices as the use of sedatives and analgesics, breaking membranes, inducing labor, episiotomy, fetal monitoring and separation of mother and child.[1–23] All of the "alternatives" suggested involve a greater degree of choice and control on the part of the parents. In particular, the use of childbirth education methods, participant father-coaches, midwives and a home or homelike environment is promoted.

A number of books have been published on these topics. The titles are indicative of the aggressive stand the authors take—*The Rights of the Pregnant Parent*,[24] *Helping Ourselves, Families, and the Human Network*,[25] *Immaculate Deception*,[26] *Birth Without Violence*[27] and *Home Birth*.[28] Whether or not these publications accurately reflect a widespread uneasiness with the interjection of sophisticated technology into the birth process is difficult to establish. Most of the women's literature is based on personal experiences and opinions or limited case studies. Its representativeness is questionable. Only a few consumer surveys of maternity care have been conducted in the United States, and their generalizability also is limited due to sampling techniques or restricted content.[29–32] These same shortcomings also contribute to some discrepancy about desired services, although all report a generally positive evaluation of care.

There are several methodological problems inherent in measuring satisfaction with maternity care. First, patients are reluctant to criticize their caretakers. This problem is intensified with maternity care, during which the arrival of a healthy baby can overshadow negative experiences and

---

[*] Assistant Professor of Sociology, Arizona State University, Tempe, Arizona.

[†] Maternity Care Nursing Consultant, Bureau of Perinatal Health, Arizona Department of Health Services, Tempe, Arizona.

Reprint requests: Deborah A. Sullivan, Ph.D., Sociology Department, Arizona State University, Tempe AZ 85281.

0025-7079/82/0300/0321/$01.00 © J. B. Lippincott Co.

create a favorable environment for subsequent evaluations of care. Nevertheless, a recent evaluation of alternative, indirect measures of satisfaction with medical care failed to show a positive association with other objective indicators of care.[33] Special attention, therefore, must be paid to the level of reported satisfaction when using a purely subjective rating rather than a rating limited to a satisfied–dissatisfied dichotomy. Second, a patient's satisfaction with care is based on his/her perception of several dimensions of caretaker conduct, including technical competence, emotional support and communication.[34] Research has shown that these dimensions are highly correlated and that a patient's perception of technical competence is influenced largely by the other measures of rapport.[35] Third, in spite of the attention given to maternity care by the popular press, many women do not perceive that alternatives could exist in their own course of care. Instead, most are dependent on caretakers to advise them, and recent ethnographic research has shown that only minimal information is given to maternity patients about options available to them.[36] British consumer surveys reveal a similar finding and observe that the desires and objections of a minority deserve special attention in this context.[37] These findings suggest that consumer surveys of maternity care need to be designed to present alternatives objectively and to probe experience with, and evaluation of, specific procedures and the caretaker–patient rapport.

This report is based on a statewide survey of consumer satisfaction with maternity care. The focus is the relationships between expressed level of satisfaction and two aspects of maternity care—communication between the woman and her caretakers and choice of specific clinical procedures. It is hypothesized that the better the rapport and the greater the fulfillment of choice, the higher will be the level of satisfaction

Methods

In the winter of 1979, the Arizona Department of Health Services conducted by mail a survey of all resident women to whom a birth certificate had been issued between October 15 and November 15, 1978. This time period avoided spurious effects in the timing of births due to major holidays and vacation schedules. A January mailing allowed time for the recording of births and the centralized collection of records from outlying areas in the state. Also, it gave the women time to complete postpartum care.

Of the 3,773 questionnaires mailed, 145 were returned undeliverable, and 1,900, or 52.4 per cent of those delivered, were completed and returned by mid-March, 1979. State confidentiality requirements did not allow a follow-up of the nonrespondents. A comparison of the characteristics of respondents with all women in the sample (Table 1) shows an underrepresentation of teenagers and women from low socioeconomic backgrounds. Separate analyses were conducted for each age, educational and ethnic category. The relationships between satisfaction and communication and choices were consistent across categories, although the overall level of satisfaction was lower for the underrepresented groups. The separate analyses are not presented in this report because of repetition and the extensive overlap among the underrepresented categories. The underrepresentation of low socioeconomic groups introduces a conservative bias in the analysis. Expressed level of satisfaction among respondents is higher than that in the population from which they are drawn.

The consumer satisfaction questionnaire included an introductory letter from the Department of Maternal and Child Health, Arizona Department of Health Services. The letter explained the purpose of the study for the Advisory Task Force on

TABLE 1. A Comparison of the Characteristics of Respondents with Women to Whom Birth Certificates Were Issued in Arizona, October 15 to November 15, 1978

	Original Sample		Respondents	
	N	Per Cent	N	Per Cent
Age				
Under 15	7	0.2	4	0.2
15–19	624	16.5	215	11.4
20–24	1,396	37.0	684	36.2
25–29	1,045	27.7	618	32.7
30–34	502	13.3	274	14.5
35–39	171	4.5	87	4.6
Over 40	28	0.7	10	0.5
Unknown	0		8	
Education				
Under 9 years	395	10.5	119	6.3
9–12 years	2,254	59.7	921	48.7
13–16 years	988	26.2	684	36.1
Over 16 years	136	3.6	169	8.9
Unknown	0		7	
Number of children born				
1	1,529	40.5	837	44.1
2–4	1,979	52.4	961	50.6
5 or more	265	7.0	98	5.2
Unknown	0		4	
Ethnic group*				
Anglo	—	72.1	1,418	74.9
Mexican-American	—	18.7	273	14.4
Black	—	2.9	46	2.4
Indian	—	5.5	110	5.8
Other	—	0.6	47	2.5
Unknown	—		6	

* Arizona birth certificates contain information on race but not ethnicity. This does not permit identification of the largest minority—Mexican-American. Therefore the per cent breakdown is based on the ethnic composition of the state in 1975.

Alternatives in Maternity Care and insured confidentiality, since neither the respondents nor their health caretakers could be identified.

The questionnaire, which underwent several pretests and revisions, consisted of 60 structured questions. The questions covered socioeconomic and demographic characteristics of the respondents, prenatal care, labor and delivery care, and postpartum care. Two sets of questions concerned satisfaction with, and quality of, communication with caretakers during the prenatal period and during labor and delivery. Another series of questions asked respondents whether they had specific clinical procedures during labor and delivery and whether or not they had wanted these procedures. In addition, respondents were asked directly for evaluations of their prenatal and labor and delivery care. The rest of the questionnaire concerned postpartum care.[38]

Contingency table analysis was used to investigate the relationship between satisfaction with care and communication patterns and clinical procedures. Associations were evaluated with chi-square tests, and the strength of the associations was measured with Cramer's V. Chi-square values having probabilites of 0.01 or less were considered significant.

Results

Respondents ranged from under 15 years of age to over 40. One quarter were part of a minority ethnic group. Family income varied from less than $5,000 to greater than $20,000. Eighty-eight per cent were married at the time of the survey, and 44 per cent were reporting on their first birth. The vast majority of respondents, 88 per cent, saw a medical doctor for prenatal care, and only 53 of the 1,900 births were delivered outside of a hospital.

Communication during Prenatal Care

The overall evaluation of prenatal care by respondents is extremely positive. Ninety-three per cent reported that they were "satisfied" or "very satisfied" with their care. The majority gave their caretaker favorable evaluations on the three questions about quality of communication during prenatal care (Table 2).

The relationship between quality of communication and overall level of satisfaction with prenatal care can be seen in Table 3. Each of the three measures of communication has a statistically significant relationship with satisfaction with prenatal care. The association is strongest for time spent discussing problems and empathy on the part of the examiner. The less time spent discussing problems and the less empathy shown by the caretaker, the lower is the level of satisfaction.

The association between vocabulary used and level of satisfaction is not as straightforward. The decline in level of satisfaction with the less frequent use of words understood by respondents is reversed for respondents who reported that their caretaker "never" used words they could understand. Perhaps these respondents regarded medical terminology as a reassuring indicator of the expertise of their caretaker. Another possibility is that the medical terminology intimidated women and created a setting in which they were expected to be satisfied regardless of their ability to understand the words used by caretakers.

Communication during Labor and Delivery Care

The findings on labor and delivery care are similar to those on prenatal care. Level of satisfaction is positively skewed (Table 4). An even larger proportion of respond-

TABLE 2. Communication Patterns during Prenatal Care and Labor and Delivery*

	N	Always %	Usually %	Occasionally %	Never %	Do Not Know %	Total %
Prenatal care							
Spent enough time discussing problems	1,887	50	31	14	4	1	100
Used words I could understand	1,880	57	24	10	9	0	100
Tried to understand how I felt	1,877	50	29	11	6	4	100
Labor and delivery							
Explained procedures	1,870	65	24	7	4	0	100
Kept me /support person informed of progress	1,860	68	19	8	4	1	100
Explained choices available	1,843	61	16	6	13	4	100

* Respondents with missing values are not included in the table.

TABLE 3. Satisfaction with Prenatal Care by Perception of Care Received*

	N	Very Satisfied %	Satisfied %	Unsatisfied %	Very Unsatisfied %
Enough time spent discussing problems					
Always	941	81	17	1	1
Usually	576	42	53	4	1
Occasionally	263	18	62	16	4
Never	78	19	28	30	23
Do not know	21	38	43	14	5
Chi square = 191.91				Cramer's V = 0.16	
Used words I could understand					
Always	1,061	70	26	2	2
Usually	449	38	51	9	2
Occasionally	189	35	52	10	2
Never	162	58	33	5	4
Do not know	12	50	25	25	0
Chi square = 62.45				Cramer's V = 0.09	
Tried to understand how I felt					
Always	943	81	18	1	1
Usually	550	45	52	3	0
Occasionally	208	14	70	15	1
Never	99	8	30	38	23
Do not know	71	39	48	11	1
Chi square = 163.11				Cramer's V = 0.15	
Total	1,885	58	35	5	2

* Respondents with missing values or those few reporting mixed satisfaction are not included in the table.

ents felt that their care takers were more responsive to their needs during labor and delivery than during the prenatal period (Table 2). However, it is during this stage of maternity care that individual experiences seem to differ greatly. Among 500 unsolicited comments on the questionnaire were some very positive endorsements, such as "Doctor . . . was fantastic," or "OB staff was excellent." Others wrote of lack of attention, insensitivity and incompetence. For example, one woman wrote that when her water broke the "nurse yelled about getting her wet!" and another wrote, ". . . the doctor was grouchy and unpleasant. He left a sponge in my vagina and didn't stitch me right."

Many of the women offering highly negative comments still checked "satisfactory" in response to the question evaluating overall labor and delivery care. This supports the previously discussed reluctance to criticize caretakers and makes the distinction between "very satisfied" and "satisfied" especially important for this study.

The strong association found between perception of care and level of expressed satisfaction with prenatal care is also evident with labor and delivery care, as seen in Table 4. Respondents who felt that their caretakers did not communicate with them about their labor and delivery experience were far more likely to be unhappy with that experience. Each of the three measures of communication has a statistically significant relationship with satisfaction as measured by a chi-square test.

TABLE 4. Satisfaction with Labor and Delivery Care by Perception of Care Received*

	N	Very Satisfied %	Satisfied %	Unsatisfied %	Very Unsatisfied %
Explained what they were going to do					
Always	1,197	64	30	3	2
Usually	437	43	48	6	3
Occasionally	135	33	45	12	10
Never	77	23	40	16	20
Do not know	8	50	38	12	0
Chi square = 221.78				Cramer's V = 0.17	
Kept me informed of progress					
Always	1,251	64	30	3	2
Usually	347	37	51	8	4
Occasionally	139	31	43	13	13
Never	80	31	38	12	19
Do not know	27	44	48	7	0
Chi square = 229.50				Cramer's V = 0.18	
Explained choices available					
Always	1,116	67	28	3	2
Usually	291	43	48	6	3
Occasionally	113	29	51	11	9
Never	230	31	45	12	11
Do not know	78	44	46	5	5
Chi square = 213.87				Cramer's V = 0.17	
Total	1,876	55	35	5	4

* Respondents with missing values are not included in the table.

Choices during Labor and Delivery

The series of questions about specific procedures during labor and delivery allow us to investigate the extent to which women have preferences concerning the management of events during labor and delivery and whether the fulfillment of such preferences is associated with the level of satisfaction. Respondents who had cesarean births were instructed to answer all of the questions that applied to them before their surgery.

It is clear from Table 5 that women do have preferences about how labor and delivery events are managed. The data show a widespread desire for a childbirth coach and encouragement to use breathing and relaxation techniques, and a desire to avoid medication that renders a woman unconscious for delivery. There is also considerable support for freedom to move around and the use of a fetal monitor. In contrast to the general uniformity of preferences about these procedures, there is a striking divergence of desire about others. These include having family members present, using sedatives and analgesics during labor, and having choice about the atmosphere during delivery.

The extent to which such preferences are honored (no discrepancy between what they wanted or did not want and what they had) varies depending on the procedure (Table 5). When such preferences are not honored, the reported level of satisfaction declines significantly (Table 6). Since few women actually rated their labor and de-

TABLE 5. Percentage Distribution of Respondents According to Procedures Wanted Versus Those Had During Labor and Delivery*

	N	Wanted		Not Wanted		Total Discrepancies
		Had	Not Had	Had	Not Had	
Childbirth coach	1,678	82	9	1	8	10
Encouragement to use breathing/relaxation	1,621	86	6	2	6	8
Freedom to move around	1,582	62	23	3	12	26
Fetal monitor	1,497	62	12	7	20	19
Family member present	1,648	43	12	2	42	14
Choice of atmosphere	1,379	23	38	2	38	40
Medication to assist me in labor	1,620	50	8	4	38	12
Medication to put me out for delivery	1,668	9	5	3	83	8

* Respondents with missing values for either "wanted" or "had"—because of a cesarean birth, because they did not choose to answer, or because they said they did not know whether they wanted or had a procedure—are excluded from the table.

livery experience negatively, the greater impact of unfulfilled preferences can be seen in the shift toward more women responding that their labor and delivery experience was "satisfactory" instead of "very satisfactory."

Discussion

Contrary to the impressions given by the popular press, our statewide survey reveals widespread satisfaction with maternity care. This finding must be tempered with the acknowledgment that satisfaction levels among our respondents are somewhat higher than those of the population because of the underrepresentation of teenagers and low socioeconomic groups. Also there is evidence that women hesitate to criticize maternity care. Even though some respondents reported that their prenatal caretaker only "occasionally" spent enough time discussing problems, using understandable words or trying to understand how they felt, the majority rated their prenatal care as satisfactory. Similarly, the majority of the relatively few women who said that their caretakers never explained what they were doing to them during labor and delivery, or kept them informed on their progress or explained the choices available to them, still said their care was satisfactory.

One explanation for positive evaluation of such unusually deficient care is the joy surrounding the arrival of a healthy baby. The healthy baby legitimizes the entire pregnancy experience and creates a favorable halo. This undoubtedly results in an upward bias in the expressed level of satisfaction. The halo also may indirectly influence satisfaction by enhancing a woman's evaluation of the quality of communication she had with her caretaker and whether a specific clinical procedure was desirable.

An additional explanation is the more subtle effect of socialization of women as

TABLE 6. Satisfaction with Labor and Delivery Care by Preferences about Specific Procedures*

	N	Very Satisfied %	Satisfied %	Unsatisfied %	Very Unsatisfied %
Childbirth Coach					
Choice honored	1,501	58	34	5	3
Choice not honored	168	41	42	10	6
Chi square = 31.24				Cramer's V = 0.14	
Encouragement to use breathing/relaxation					
Choice honored	1,483	60	33	5	3
Choice not honored	130	35	46	10	9
Chi square = 41.86				Cramer's V = 0.16	
Freedom to move around					
Choice honored	1,169	64	30	4	2
Choice not honored	406	41	44	8	8
Chi square = 81.29				Cramer's V = 0.23	
Fetal monitor					
Choice honored	1,216	60	32	5	3
Choice not honored	274	44	41	8	7
Chi square = 23.92				Cramer's V = 0.13	
Family member present during labor					
Choice honored	1,398	59	33	5	3
Choice not honored	239	41	45	8	6
Chi square = 29.14				Cramer's V = 0.13	
Atmosphere					
Choice honored	830	65	30	4	2
Choice not honored	545	46	40	8	6
Chi square = 58.98				Cramer's V = 0.21	
Medication to assist me to relax in labor					
Choice honored	1,426	61	33	4	2
Choice not honored	188	32	43	12	13
Chi square = 106.27				Cramer's V = 0.26	
Medication to put me out for delivery					
Choice honored	1,537	59	34	4	3
Choice not honored	123	27	41	16	16
Chi square = 11.99				Cramer's V = 0.29	

* Respondents with missing values are not included in the percentages.

potential maternity patients. The rapid de-[illegible]nes in maternal mortality during the [illegible] reated a social consensus in the [illegible] about the role of medical practitioners as interventionistic caretakers and the role of women as passive recipients of care. This duality brought with it the potential for communication problems and loss of decision making, which most women accept, or at least tolerate, as part of the modern science of obstetrics.

At a time when the popular media are presenting numerous challenges to standard obstetric care, the level of satisfaction with maternity care is an important factor in the future choice of caretaker and type of care. It was concern about the increasing number of home births in the state that led to the present study. Many of the 39 respondents who had a planned home birth attached letters to their questionnaires explaining their choice. Most had previously borne children in a hospital setting. Few related any unusually negative experiences. The most common theme, echoing the popular media, was a desire for a more personal experience. The women described this desire in terms of greater rapport with their caretaker and more flexibility in procedures, which allowed them more control over their birthing experience. Some of those using the services of midwives in a hospital setting made similar comments.

The results of our study suggest that the desire for rapport and flexibility are not confined to the extremely small number of women resorting to radically different forms of maternity care. Indeed, for the much greater proportion of women who had just completed a more traditional obstetric experience, extent of rapport and fulfillment of clinical preferences were strongly associated with their level of satisfaction. These associations may be inflated to some extent due to the halo effect previously mentioned. Nevertheless, these findings give tentative indication that users of standard obstetric systems of care do not differ greatly from users of radically different systems in terms of the type of relationship they want with their caretakers. They do differ somewhat in their choices of clinical procedures.

Additional analyses, not included in this report, indicate that users of alternative care systems uniformly wanted to avoid medication and other interventionistic practices. In contrast, users of more standard obstetric care systems expressed much greater variation in desire on some of the clinical procedures, especially having family members present, use of medication in labor and choice of atmosphere for childbirth. Some of the variation in expressed desires may be the result of subsequent rationalization of experiences that resulted in a healthy baby.

The data show that the popular women's press does accurately reflect the interests of the majority of women on some clinical procedures and of a sizable minority on other issues. Even allowing for rationalization of experiences, there is much desire for childbirth coaches and controlled breathing and relaxation, and little desire to be rendered unconscious for delivery. A majority still wanted medication during labor in addition to breathing and relaxation techniques, but a large minority, 41 per cent, preferred none. On these four issues, most women's desires were fulfilled. When they were not, level of expressed satisfaction was significantly lower. Of all the preferences, those surrounding uses of medication seem most salient for level of satisfaction.

Many women's choices were not honored on two other alternatives, the freedom to move around and choice of atmosphere (defined on our questionnaire as "dim lights, quiet music, warm bath, pictures, etc."). Taken together with the preference about the presence of family members, the data suggest support for the homelike birthing room alternative advocated in the women's media among more than half the respondents. One hospital in the state that has implemented a birthing room program reports success in averting some home births and enhancing consumer satisfaction.[39]

Implications

Because of the low levels of maternal and infant mortality achieved by modern obstetric practices, it might seem gratuitous to be concerned with consumer satisfaction with maternity care. If satisfaction were based solely on outcome, virtually all consumers would be "very satisfied" with the care they received. The fact that all consumers are not points to the importance of the experience as well as the outcome.

The uneasiness that some consumers have felt with their maternity experience has been transformed into a small but vocal movement seeking to modify current practices. Few reject modern obstetric practices, although they argue for greater choice and control for the parents. Their expressed aim is to personalize maternity care.

The results of this survey suggest that a more personal, more satisfactory course of care depends on more attention to the interpersonal relationships between caretakers and patients and greater flexibility in clinical procedures. To a large extent these two suggestions are intertwined. If caretakers work to establish a good rapport with their patients, they will be advocates for their patients' desires concerning the management of labor and delivery. Similarly, if patients have a good rapport with their caretakers, they will understand the desirability of recommended procedures. Such a multifaceted approach would increase the efficiency of care rather than create more problems. At the same time it would remove some of the motive that has led a small but increasing number of prospective parents to chance less safe procedures.

References

1. Albertson M. Seven choices of delivery. Good Housekeeping 1978;186:68.

2. Ander M. Return of the midwife. Good Housekeeping 1976;183:198.

3. Bing E. Childbirth: liberated labor. Harper's Bazaar 1976;110:103.

4. Brazelton TB. Welcoming your baby: natural childbirth. Redbook 1979;153:35.

5. Clark NH. Traditional alternatives: Lamaze method at maternity wing of New York Hospital Cornell Medical Center. Harper's Bazaar 1976;110:103.

6. Davis F. Who's having the baby—you or the doctor? Woman's Day 1976;March:10.

7. Frank A, Frank S. Home or hospital delivery. Mademoiselle 1978;84:66.

8. Gardner J. Having your baby at home. Glamour 1979;77:240.

9. Hoover MB. Hidden rewards of childbirth training. McCall's 1974;101:59.

10. Lake A. Childbirth in America. McCall's 1976;103:83.

11. Marshall MC. My husband delivered our baby. Good Housekeeping 1976;179:77.

12. Maynard F. Home births vs. hospital births. Women's Day 1977;June:10.

13. Milinaire C, LeBoyer F. Born happy. Vogue 1974;164:80.

14. Norwood C. Humanizing way to have a baby: with nurse midwives at birth centers. MS 1978;6:89.

15. Panter GG. Prepared childbirth—moving ahead. Parents 1979;54:33.

16. Pascoe ES. How women want to have their babies. McCall's 1977;105:109.

17. Powers T. Fathers in the delivery room. Good Housekeeping 1974;179:93.

18. Randal J. Where should your child be born? Parents 1979;54:74.

19. Rose L. Pregnant women's guide to pain relief for labor and delivery. Glamour 1980;78:152.

20. Samuels M, Samuels N. Relax and visualize your way to a better pregnancy and birth. Parents 1979;54:48.

21. Torres T. From womb to world: a nonviolent transition: the LeBoyer method. MS 1974;3:22.

22. Trezise J. Midwives move into mainstream. Working Woman 1978;3:63.

23. Yuncker B. Delivery procedures that endanger a baby's life. Good Housekeeping 1975;180:56.

24. Elkins V. The rights of the pregnant patient. New York: Two Continents Press, 1976.

25. Howell M. Helping ourselves, families and the human network. Boston: Beacon Press, 1975.

26. Arms S. Immaculate deception. Boston: Houghton-Mifflin, 1975.

27. LeBoyer F. Birth without violence. New York: Random House, 1976.

28. Petty R, ed. Home birth: a complete, authorative guide on the back to basics trend in childbirth. Chicago: Domus Books, 1979.

29. Nunnally D, Agular M. Patients' evaluation of their prenatal and delivery care. Nurs Res 1974;23:469.

30. Scaer R, Korte D. MOM survey: maternity options for mothers—what do women want in maternity care. Birth Fam J 1978;5:20.

31. Anderson SF, Kay MA, Shaw E.. Childbirth as they like it. Arizona Nurse 1978;May/June:4.

32. Norr KL, Block CR, Charles A. Explaining pain and enjoyment in childbirth. J Health Soc Behav 1977;18:260.

33. Steward M, Wanklin J. Direct and indirect measures of patient satisfaction with physicians' services. J Community Health 1978;3:195.

34. Ware JE, Doyle BS. Physician conduct and other factors that affect consumer satisfaction with medical care. Publication No. D-5670. Santa Monica: Rand Corporation, 1977.

35. DiMatteo MR, Hays R. The significance of patients' perceptions of physician conduct. J Community Health 1980;6:18.

36. Riley E. What do women want? The question of choice on the conduct of labor. In: Chard T, Richards M. Benefits and hazards of the new obstetrics. Philadelphia: JB Lippincott Co, 1977.

37. Danziger SK. Treatment of women in childbirth: implications for family beginnings. Am J Public Health 1979;69:900.

38. Sullivan DA, Beeman R. Satisfaction with post-partum care: opportunities for bonding, reconstructing the birth and instruction. Birth Fam J 1981;8:153.

39. Kieffer MJ. The birthing room concept at Phoenix Memorial Hospital: consumer satisfaction during one year. JOGN Nurs 1980;9:151.

MEDICAL CARE
March 1982, Vol. XX, No. 3

Utilization of Specialized Ambulatory Care by the Elderly

A Study of a Clinic

STANLEY R. INGMAN, PH.D.,* AND IAN R. LAWSON, M.D.†

Specialized ambulatory geriatric care remains an understudied topic. This article analyzes the organization of one special clinic that was established in an apartment block of elderly housing. The findings indicate that as the number of nurse practitioners and case workers' visits rose steadily during the first four years of the clinic's existence, visits to the clinic physicians declined by approximately one visit per client per year. The clinic population utilized less inpatient hospital care than other nonusing residents, and less than other residents of similar age in Connecticut. There was reduced usage of intermediate nursing home care (as opposed to skilled nursing care). Hence, how geriatric ambulatory care is delivered affects institutionalization. In the interests of general well-being and economy, more attention should be paid to this extramural phase of medical care of the elderly.

CRITICS OF THE HEALTH CARE SYSTEM for older adults have charged that it is biased toward institutionalization, whether it be in the hospital or in the nursing home.[1,2] With the degree of sickness and disability prevalent among the elderly, alleviation of disability and prevention of its more severe, dependency-inducing forms may be expected to have effects on institutional admission rates. There has been marked interest in home care services and improved socioclinical assessment as part of community, state and federal strategies to decrease institutionalization.[3] Ambulatory geriatric care has not received the same attention, yet may have similar potential for reducing the utilization of institutional services.[4,5]

Medicare has concentrated federal and private insurance reimbursement on acute hospital care, as Medicaid has supported the expansion of nursing home care. National data indicate that those aged 65 years and over now occupy slightly less than 40 per cent of the hospital beds on any one day. From 1967 to 1975 there has been a 61 per cent increase in hospital admissions for those aged 65 years and over, as opposed to only 16 per cent for those under 65 years of age.[6]

This increase in the utilization of hospital services by the elderly is probably the result of two major processes: 1) the increased availability of these services to the elderly because of Medicare and 2) their financial advantage to the using physician.

From the Department of Family and Community Medicine, University of Missouri-Columbia, School of Medicine,* and the Department of Community Medicine and Health Care, University of Connecticut, Storrs, Connecticut.†

Supported in part by the Community Studies Unit, Department of Community Medicine and Health Care, UCONN-Health Center.

Dedicated to Sister Barbara Freeman, Nurse Practitioner in the program, who died September 9, 1980. The authors appreciate the careful comments and suggestions from Derek Gill, Ph.D., and Mike Stotts, M.S. Leonard Menczer, D.D.S., M.P.H., Hartford City Health Department, was instrumental in the program's initiation and growth.

Reprint requests: Stanley R. Ingman, M-330A, University of Missouri-Columbia, School of Medicine, Columbia, MO 65201.

0025-7079/82/0300/0331/$00.90 © J. B. Lippincott Co.

Studies of Ambulatory Care for the Aged

Wan in 1977 reported a study of 2,000 noninstitutionalized older persons living in low-income urban areas near five different neighborhood health centers. Those patients in the catchment areas surrounding the neighborhood health centers increased their utilization of ambulatory medical care by a factor of three as compared with a similar group of elderly persons outside the area covered by these facilities.[7]

In Baltimore, German et al. point out that the type and nature of ambulatory care delivery does increase the likelihood that elderly individuals will request care for less life-threatening but potentially serious conditions. In addition, the elderly using the health maintenance organizations in Baltimore are more likely to be seen for symptoms than were those receiving care through the traditional system of medical care, e.g., physician office practice or hospital outpatient departments.[8]

Davis and Reynolds (1976), using the National Health Survey data, showed that health status is the most important predictor of physician and hospital utilization and that those that are ill use more services. This challenges the commonly held opinion that if services are made "too" accessible, the "well-lonely" elderly will overutilize those services.[9]

Barney and Neukom argue that the elderly need not only more services, but that they need specialized health services.[10,11] In Michigan they observed that many elderly seem to be rather circumspect and pessimistic about conventional health care, i.e., they commonly hold the view that "efforts to ameliorate the symptoms of decline involve excessive cost for questionable benefits.[11] Such health beliefs (adopted by many health care providers as well as elderly) according to Barney and Neukom (1979), serve to explain the national problem of meager provisions and meager use of geriatric care. Barney and Newkom describe the main points of that belief system: "old age is a malady in which all are susceptible and all will succumb, in treating its symptoms health care is increasingly costly and increasingly ineffective."[10] Several actions to combat these beliefs and improve the provisions of geriatric services are outlined by Barney and Neukom.[11]

> First, because it is not clearly known that health services are of true benefit to the older patients, more programs like the Pilot Geriatric Arthritis Program are needed to carry on demonstrations and evaluations. Second, on the basis of what is known, policy planners should allocate resources to health services for the elderly, without fear that the patient demand will get out of control. Finally, policy planners should consistently allocate a portion of the geriatric health care resources, as did the Arthritis Program, to designing, publicizing, and delivering services, in such a way as to reach the isolated, the needy, and the reluctant elderly patient.

By these means, the authors feel that geriatric health care could play a vital role in assuring health and independence to elderly Americans.

Development of specialized geriatric health care services may be supported from the perspective of continuity of service provision. While continuity of care is considered to be one of the major advantages to be derived from the provision of specialist primary patient care, continuity of care is even more apposite for the care of the elderly. Here, the mix of social and medical factors in illness are complex, and continuity in the provider system is essential for both understanding and management. In this article we examine a population of elderly who receive ambulatory care service where there were no out-of-pocket costs in terms of providing health maintenance, continuity of care, a range of social and clinical support services, and "routine" medical care. The description concentrates on the quality and type of medical care utilized by the participants in the Smith Tower Clinic.

The Geriatric Medical Clinic

In 1972 at the invitation of the Hartford City Health Department, the clinic was established in a 200-apartment building for low-income elderly. The Housing Authority was permitted to allocate two apartments to this clinic and to a podiatry/dental clinic. A secretary, a nurse practitioner and a case coordinator were the full-time staff five days a week. Typically, physicians' sessions were held three mornings a week. Medical and social histories based on a protocol were taken by the nurse practitioner, either in the patients' homes or on their first visit to the clinic. Physical examination was also conducted by protocol.

Elderly residents in the apartment building and in the community were told that attendance at the clinic required a commitment to the clinic program, that is, they were to view the clinic as the sole service for primary care and specialty referral.

After a comprehensive history and examination were completed by the nurse practitioner, every patient was seen by an involved physician who made selective follow-up examinations. Each visit to the nurse practitioner case coordinator and physician was recorded in the clinical record, as were admissions to other facilities in Hartford. The record format was problem oriented, with taxonomic ordering[12] of the master problem list; drugs being taken, target symptoms and selected laboratory results were listed on a flow sheet. Up-to-date copies of these were available at the related hospital's emergency rooms and accompanied patients when they were referred out to specialty offices and clinics.‡

After three months as a free-standing clinic, the clinic was incorporated into the Mt. Sinai Hospital outpatient department. This allowed the clinic to receive reimbursement from Medicare and Medicaid and to ensure better integration with hospital emergency services. The city continues to contribute funds to permit the hospital to waive Medicare co-insurance and deductible charges for all elderly patients. Patients therefore meet only the costs of medication. Private specialty consultants agreed to accept assignment and frequently waived billing for co-insurance and deductible. The Hartford Foundation for Public Giving provided basic office equipment.

Clinical records form the basis of the analysis presented in this article. In 1976 the records from 1972–1975 were coded and processed for computer analysis. The records included data derived from the first visits, together with information presented at subsequent visits. More research funding was secured in 1977, so a similar process of coding was done for patients entering the clinic in 1976 and 1977. Therefore, when analyzing certain outcome factors such as present status, we will separate the two clinic populations in our analysis (i.e., 1972–1975 group and 1976–1977 group).

A prospective comparative study of the aged in various high-rise apartments was being done by Dr. Adrian Ostfeldt of the Yale School of Public Health during the 1970s. Since some of his subjects became patients at our clinic, we were able to compare how their health fared in relation to that of others not served by our clinic program.

Characteristics of Clinic Users

Between May 1972 and December 1977, approximately 644 patients attended the clinic. Two thirds of the clients were women; 49 per cent of the clients were between 71 and 80 years of age; 21 per cent were 81 years old or older; and 30 per cent were under the age of 71 years. Although the clinic was advertised as a geriatric clinic, 13 people under the age of 60 years were attracted or referred to the clinic. Approximately 17 per cent of those entering the clinic between 1972 and 1977 were black elderly. With respect to living arrangements, 65 per cent lived alone upon entry into the clinic, and 25 per cent were married. Between 1972 and 1977, 82 per cent of the clients seen reported a monthly income of $300 or less. Slightly less than 10 per cent of the clients reported receiving $600 or more per month.

---

‡ Full operational description of the clinic and the nurse practitioner role is available from the authors.

In the initial years of operation, more than 80 per cent of the clients lived at the Smith Tower apartment building (73 per cent) or within four blocks of the Towers (7 per cent). By 1975, more than 80 per cent of the new patients entering the program lived outside the four-block area. In 1976 and 1977, of the 150 new patients entering the program, only 27 came from within the immediate four-block area. By December 1977, five years after the introduction of the clinic, of the 250 residents in the Smith Towers, 153 had attended the clinic; 96 of this group began attending in 1972.

Between 1972 and 1977 the proportion of those newly registering (a low of 20 per cent in 1973 and a high of 72 per cent in 1977) stated that they previously had a private physician. Nevertheless, throughout the six-year period, only 50 per cent of the new registrants reported seeing a physician in the preceding six months, and 12 per cent stated that they had not seen a physician for more than three years. It would seem that as the clinic became better known to the tenants of Smith Towers and the surrounding area, patients were attracted to the clinic even when they already had a private physician.

Thirty-six per cent had seen a professional about glasses in the year prior to their joining the clinic. Most had not visited a podiatrist (65 per cent), and only 18 per cent had been to a podiatrist in the previous year. Only 25 per cent had been to a dentist in the year prior to their joining the clinic.

Health Status of Clinic Users

Population at Intake. The incidence of selected disabilities of daily living at intake is presented in Table 1. These disabilities, the incidence of particular symptoms, and use of hypnotics and stimulants (see Table 2) are derived from direct leading questions per protocol.

This high incidence of symptoms is commensurate with half of the new registrants in each year declaring themselves to be in only "fair" or "poor" health.[13,14]

To summarize, the clinic population consisted of people living alone with low incomes; as a group, while ambulatory, they carried a high "load of morbidity." In time the clinic attracted more people living farther from the clinic each year and more people who had a regular private physician prior to attending the clinic.

Utilization of Clinic Services

Over four years (1972–1975) of the clinic's operation, as shown in Table 3, there was a decline in visits to the physicians—from an average of 4.0 visits per patient in 1972 to an average of 3.3 visits per patients in 1975.§ In contrast, visits to the nurse practitioners increased every year. In 1972, there was an average of 2.9 visits per year to the nurse practitioner, and an average of 4.8 visits per year in 1975. Twenty-seven per cent of patients in 1972 and 60 per cent in 1975 visited the case coordinator. Those visiting the case coordinator at least once averaged 1.3 visits from 1972 to 1974. In 1975, this rate rose to 5.3 visits per patient to the case coordinator.

Thus, in the initial four years of the clinic's operation there was a decline in physician utilization. In addition, from 20 per cent to 26 per cent of the active patients

---

§ The National Center for Health Statistics reported that there were on average 4.9 visits to a physician's office by persons 65 years and older (NIA, 1976).

TABLE 1. Levels of Disability Upon Entry (N = 644)

Activities of Daily Living	1972-1977
Difficulty walking up or down stairs	37%
Difficulty cutting toe nails	35%
Difficulty walking	31%
Difficulty washing/bathing	13%
Help to go to toilet needed	4%

(1972–1975) each year did not choose to visit or were not asked to visit one of the clinic physicians. For those patients who see the nurse practitioner and case coordinator, the average number of visits increased steadily until 1976.

In the 1976–1977 cohort, the utilization rates for the 1976 intake (N = 82) were 2.9 physician visits, 3.1 nurse practitioner visits and 3.5 case coordinator visits per year.

In 1977, the average number of visits for patients who entered in 1976 as well as in 1977 was 3.5 visits to the physician at the clinic. In contrast to 1973, there were fewer visits on the average to the nurse practitioner in 1977 ($\overline{X}$ = 2.4). Case coordinator visits also dropped off to an average of 1.2 per patient (range = 1–20) in 1977, as did visits to the physicians at the clinic in 1977. One possible explanation for the decline in clinic utilization relates to pressure from third-party intermediaries for the clinic *not* to see residents for health maintenance or chronic disease follow-up visits.

TABLE 2. Single Symptoms and Selected Medications Taken at Time of Assessment and Entering Clinic (N = 644)

Single Symptom/Drug Taken	1972-1977
Tired	58%
Poor sight	55%
Dental problems	46%
Shortness of breath	44%
Poor sleep	42%
Taking sleeping pills	19%
Use of alcohol	40%
Use of tobacco	39%
Constipation	39%
Taking laxatives	36%
Deafness	35%
Hearing aid worn	5%
Leg swelling	31%
Falls	28%
Unsteadiness on feet	28%
Skin problems	27%
Weight loss	25%
Poor appetite	23%

Many factors influence the increase or decrease of visits to specific occupational groups in the clinic. For example, clinic and insurance company policy, as well as patient behavior, are only three factors. However, the crucial issue is to assess the impact of different visitation patterns on the utilization of other services and on general well-being.

Utilization of Institutional Services by Clinic Participants

Based on the recollection of the patients, the average percentage hospitalized for the year prior to joining the Smith Tower Clinic was 17 per cent from 1972 to 1977 (see Table 4). From clinic records in 1972, it was determined that 47 per cent of those clients who were active in the clinic in 1972 were admitted to a community hospital in Hartford. Of those patients who entered in 1972 and who were still clinic attenders in 1975, 22 per cent were admitted to a hospital. Higher levels of morbidity of early cohorts may explain the initially high hospitalization rates of 47 per cent and 30 per cent. Even with increased access to ambulatory care at the Smith Tower clinic, the percentage of patients hospitalized did not rise, but declined in time. The percentage of patients hospitalized in Connecticut for all persons aged 65 and older was 26 per cent in 1976.[15] In general, the clinic population seemed to utilize hospital services significantly less, both before and after joining the clinic program, than the typical older adult in Connecticut.

On the issue of the average length of stay in the hospital, statistics fluctuated from year to year. The average length of stay per

TABLE 3. Average Number of Visits to Smith Tower Clinic: Physicians, Nurse Practitioners and Case Workers*

Number of Active Patients by Year	Number of Visits to Physicians	Number of Visits to Nurse Practitioners	Number of Visits to Case Workers
1972–1975 Cohorts			
N = 133/1972	4.0	2.9	1.7
N = 231/1973	4.5	4.0	.9
N = 369/1974	3.5	4.6	1.5
N = 410/1975	3.3	4.8	5.3
1976–1977 Cohorts			
N = 82/1976	2.9	3.1	2.5
N = 129/1977	3.5	2.4	1.2

* Calculation includes residents who may not have visited a physician, nurse practitioner or case worker in a specific year.

admission per year ranged from 12.2 to 16.9 days (1972–1975). For the 1976–1977 cohorts, the average length of stay for active patients was 13.0 in 1976 and 11.0 in 1977. The average length of stay in Connecticut during this time was 11.9 days for persons admitted who were aged 65 or older (CHA Reports, 1978). This may suggest that clinic patients who did require admission to a hospital were more ill than the typical Connecticut older adult admitted to a hospital, and thus spent more time when admitted.

TABLE 4. Hospitalization of Clinic Patients

Number of Active Patients Per Year	Percentage Hospitalization	Average Length of Stay
1972 (N = 133)	47	12.2
1973 (N = 231)	30	13.8
1974 (N = 369)	22	16.6
1975 (N = 410)	22	16.1
1976–1977 Cohorts		
1976 (N = 82)	20	13
1977 (N = 129)	18	11

With respect to nursing home use, on the average approximately 2 per cent were admitted each year (1972–1977). In the 1976 cohort, two patients were admitted out of 82 patients (2.4 per cent). In 1977, only three patients were admitted to a long-term care facility out of the 134 active patients who entered the clinic in 1976 and 1977. It is important to note that all nursing home admissions were made to skilled nursing facilities (SNF), and not to facilities providing intermediate-level care. By fully utilizing medical and social services, patients were in considerable dependency before it was thought that long-term institutionalization was warranted. In other circumstances they might have required earlier institutionalization, but to a lower (intermediate) care level.

In 1972, as previously noted, Dr. Adrian Ostfeld was doing a study of the effects of relocation of the elderly. He collected four samples of 50 elderly each in Connecticut. Of the 50 cases collected at the Smith Tower, 30 were nonattenders at the clinic and 20 were attenders. Two years later, in 1974, he noted the following: In the 30 nonattenders, 14 had been hospitalized at least once, two were admitted to Skilled Nursing Care, and there were six deaths. However, in the 20 attenders, only five had been hospitalized, two had been admitted to Skilled Nursing Care and, most striking, there were no deaths in this group. In reviewing the health status of each group using standard measures, he found no striking differences between the two groups in 1972.

Status of Patients as of June 1976

Approximately 53 per cent of those who entered the clinic between 1972 and 1975 were still active patients as of June 1976. Table 5 indicates that approximately 6 per cent died and that 5 per cent were residing in an SNF as of June 1976. Eighteen per cent chose to seek another source of primary care. The status of 10 per cent was unknown as of June 1976.

The most significant cohort for review as of June 1976 comprised those patients who had entered the clinic in 1972. Of 131 patients, as of June 1976, 51 per cent (67) were still active in the clinic; 12 per cent had chosen to go elsewhere for primary care; 16 per cent had moved out of Hartford or their status was unknown; 9 per cent had died during the four-year period; and 10 per cent were in skilled nursing facilities.

Conclusion

As Wan[7] and German et al.[8] also documented, better outpatient service accessibility tends to increase visitations. The high levels of morbidity and disability among the Smith Tower clinic population should help dispel the notion that the "worried well" will dominate geriatric ambulatory care when services are made freely accessible.

With patients seeing the nurse practitioner, the case coordinator and the physician on different visits to the clinic, the average number of visits to the Smith Tower Clinic was well above the average of 4.9 physician visits per older adult in New England (1973–1974). At the Smith Tower Clinic there was a steady decline in physician visits and an increase in the less costly nurse practitioner and case coordinator visits. As Quinn[16] suggested, when comprehensive geriatric assessment and socioclinical support services are provided to the elderly, less physician services and hospital care are required by the older adult.

Our study investigated the relationship between aggressive, comprehensive ambulatory geriatric care and hospital utilization. Our review of hospitalization prior to entry into the clinic shows that the clients had a lower average rate of hospitalization than that found within the state of Connecticut in 1976. Since poor people are more likely to avoid medical services when copayment is required, this low rate of hospitalization may not be surprising given the large number of poor elderly in our clinic population. At least this may explain the low rates prior to their joining the

TABLE 5. Status of Patients as of June 1976 by Cohorts (1972–1975)*

	1972	1973	1974	1975	Total
Active	67 (51.1%)	68 (61.8%)	90 (58.9%)	54 (53%)	279 (53%)
Moved	7 (5.3%)	6 (5.4%)	5 (3.3%)	6 (6.7%)	24 (5.5%)
Died	12 (9.2%)	5 (4.5%)	5 (3.3%)	3 (3.4%)	25 (5.7%)
SNF	13 (9.9%)	5 (4.5%)	2 (1.3%)	3 (3.4%)	23 (5.3%)
Changed primary care source	16 (12.2%)	20 (18.2%)	33 (21.6%)	12 (13.5%)	81 (18.5%)
Referred to other physician	2 (1.5%)	1 (.9%)	2 (1.3%)	2 (2.3%)	7 (1.6%)
Unknown	14 (10.7%)	5 (4.5%)	16 (10.5%)	9 (10.1%)	44 (10.1%)

The 1976 and 1977 cohorts were not included for this analysis.

clinic. However, clinic membership did increase ambulatory visits when copayments were covered by grant funds and did not cause a rise in hospitalization, as national trends would have suggested. If one notes that hospitalization for the age group nationally was increasing by approximately 5 per cent per year from 1972 to 1975, it is even more significant that the rate for the clinic population did not rise each year.[6] It is possible that intensive ambulatory care and coordination of other services secured for the attenders were able to prevent some hospitalization. Our higher than average "length of stay" for patients hospitalized, and the use of skilled nursing beds as opposed to intermediate care beds, suggest that institutional care was reserved for more dependent or ill patients.

As various community models that focus on comprehensive case assessment and management are being systematically developed and evaluated nationally, the type of geriatric ambulatory care clinic that we described in this article should also be carefully investigated. Current federal policy to establish special services for the frail elderly and to prevent inappropriate institutional admissions should deal with both outpatient care and home care. We would argue that aggressive and more broadly conceived ambulatory care systems can respond to some of the basic needs of the frail and not so frail older adults. Health care finance administrators, at a minimum, should review the Medicare administrative policy of discouraging preventive outpatient care as has been experienced by the staff when third-party intermediaries visited the Smith Tower Clinic.

References

1. Somers AR. The geriatric imperative: a major challenge to the health professions. In: Somers AR, Fabian DR, eds. The Geriatric Imperative. New York: Appleton-Century-Crofts, 1981.

2. Brody SJ. Comprehensive health care for the elderly: an analysis. Gerontologist 1978;13:412.

3. The T.G. interview: Robert Benedict. Gerontologist 1980;20:131.

4. Somers AR. The high cost of health care for the elderly: diagnosis, prognosis, and some suggestions for therapy. J Health Polit Policy Law 1978;3:163.

5. Bellin SS, et al. Impact of ambulatory-health-care services on the demand for hospital beds. N Engl J Med 1969;28:808.

6. Utilization data for senior citizens. Hospitals 1976;50:47.

7. Wan T. Access to care, health status and health service utilization of non-institutionalized older persons in low-income urban areas. Paper presented at Gerontological Society meeting, San Francisco, November 18–22, 1977.

8. German PS, et al. Health care of the elderly in medically disadvantaged populations. Gerontologist 1978;18:547.

9. Davis K, Reynolds R. The impact of medicare and medicaid on access to health care. In: Rosett RN, ed. The role of health insurance in the health service sector. New York: National Bureau of Economic Research, 1976.

10. Barney JL, Neukom JE. Elderly users of health services. Paper presented at Gerontological Society meeting, San Francisco, November 18–22, 1977.

11. Barney JL, Neukom JE. Use of arthritis care by the elderly. Gerontologist 1979;19:548.

12. Lawson IR. A taxonomy of geriatric care using the Problem-Oriented Record System. In: Lawson I, Ingman S, eds. The language of geriatric care: implication for professional review. New Haven: Connecticut Health Services Research Series, 1975.

13. Maddox GL, Douglas EB. Self assessment of health: a longitudinal study of elderly subjects. J Health Soc Behav 1973;14:87.

14. Linn BS, et al. The very old patient in ambulatory care. Med Care 1978;609.

15. Impacts of an aging population on utilization and bed needs of Connecticut hospitals. CHA Reports 1978;1:1.

16. Quinn JL. Triage-coordinated care for the elderly. J Continuing Education in Nursing 1979;10:7.

C5170-1 5/3/82

Medical Care

April 1982
Vol. 20 No. 4
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Continuity Reexamined: Differential Impact on Satisfaction with Medical Care for Disabled and Normal Children

Distress Syndromes, Illness Behavior, Access to Care and Medical Utilization in a Defined Population

Psychological Distress and Patient Satisfaction

Effects of Physical and Mental Health on Health-State Preferences

Life-Styles, Health Status and Medical Care Utilization

Stress Potential and Esotericity of Health Problems: The Significance of the Physician's Affective Behavior

Patient Suffering and Patient Satisfaction Among the Chronically Ill

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

MEDICAL CARE
April 1982, Vol. XX, No. 4

Original Articles

Continuity Reexamined:

Differential Impact on Satisfaction with Medical Care for Disabled and Normal Children

NAOMI BRESLAU, PH.D.*

The satisfaction with medical care of families with disabled children and of randomly selected families is compared to test the hypothesis that the impact of continuity of physician on parental satisfaction is greater when disabled children are patients than is the case among the general population of children. A multiple regression approach for the study of interaction was used. The results provide general confirmation for the hypothesis that the weights of continuity in producing change in all three patient satisfaction areas studied were greater in the sample of families with disabled children than in the random sample of families. Controlling for maternal education, family income, race and waiting time did not alter these results. It is suggested that a continuous doctor–patient relationship, conducive to the expression and resolution of psychological needs, is of special importance when a patient's illness is severe. Implications for the organization of medical care in specialty clinics, which serve the disabled and chronically ill, are discussed.

CONTINUOUS CARE from a single physician has generally been viewed as a distinctive prerequisite in primary care.[1–3] It is assumed that continuity promotes a stable doctor–patient relationship, and enhances patient trust in medical recommendations and physician concern and efficacy. Visits to specialists for the diagnosis and treatment of complicated medical problems are viewed as necessary interruptions of this relationship. Medical care from specialists is said to be problem focused, and with the resolution of the problem it is assumed that patients would return to their primary physicians, with whom their continuous relationship is expected to span long periods, during both illness and health.

Since studies of medical programs offering continuous care have been conducted largely in primary care settings,[4–10] the resultant literature implies that continuity of care is only relevant to primary care settings. This is consistent with the idea that primary and specialty care perform differentiated and complementary functions for which different structures are required, i.e., continuity for the functions of primary care, not for those of specialty care.

In recent years there has been a growing recognition that medical care in the United

---

Reprint requests: Dr. Naomi Breslau, Department of Epidemiology and Community Health, Case Western Reserve University, School of Medicine, 2119 Abington Road, Cleveland, OH 44106.

The author wishes to thank Judah Rosenblatt, Department of Biometry, and Marie R. Haug, Department of Sociology, Case Western Reserve University, for instructive comments on the analysis of these data. This study was supported by grants from the Cleveland Foundation and the C. S. Mott Foundation. Presented in part at the 1981 Annual Meeting of the Ambulatory Pediatric Association, San Francisco California.

0025-7079/82/0400/0347/$01.20 © J. B. Lippincott Co.

States does not fully conform to this model: the actual functions performed by primary care physicians and specialists are not as distinct as one would expect.[11] A recent analysis[12] estimated that approximately one fifth of the population receives regular ongoing care (i.e., primary care) from specialists. This fraction consists chiefly of chronically ill and disabled patients. A considerable proportion of such specialty medical care is based in hospitals and medical centers, and is organized by pooling fragments of work schedules of several practitioners whose total work effort is divided among patient care, teaching, research and administration. In contrast to the traditional small office practice, in which continuity of physician follows from the patient's continuous use of the setting, repeated use of the large medical center or hospital clinic does not in itself secure continuity of physician.[7] Although some specialty clinics have deliberately set up conditions for continuous patient care from a single physician, the typical organizational structure of the clinic produces medical care that has been described as fragmentary and discontinuous.[6] Yet, the prevailing conceptual model of medical care, which assumes a functional differentiation between primary and specialty medicine, has excluded specialty medical settings from the realm of concern with continuity of physician. And, since specialists provide ongoing care to chronically ill and disabled patients, the continuity of their care has not been a focus of concern either.

Some writers have recently suggested that continuity of physician may be more important in the care of chronically ill patients than in that of the general population.[2,3,7,13] The rationale for this suggestion has not been explicitly stated. While the requirement of consistency in the information given to severely disabled patients and other specific arguments might be advanced in support of this hypothesis, a more self-evident argument is that the characteristics of medical care, including its continuity, would have a greater influence on the attitudes of chronically ill patients who use medical care frequently than on the attitudes of those free of illness who use it more sparingly.

In a recent paper[14] we reported the effect of continuity of physician on satisfaction with medical care in a study of families of disabled children receiving care in four specialty clinics. Continuity of physician accounted for a large part of the difference in satisfaction across the clinics. Further, continuity of care contributed to patient satisfaction in specialty clinics that provided all of a patient's medical services as well as in clinics that provided only specialized care, excluding the treatment of intercurrent illness and immunizations. The results indicated that continuity of physician exerted a strong influence on the ways in which families of disabled children evaluated their doctors and the medical care they received in specialty clinics. They could not, however, answer the question of whether or not the impact of continuity on satisfaction in this sample was greater than it would be in a sample of families with children free of disabilities.

This article considers that question. The satisfaction with medical care of families with disabled children and of randomly selected families is compared to test the hypothesis that the impact of continuity on parental satisfaction is greater when disabled children are patients than is the case among the general population of children.

Measures of satisfaction with medical care constitute valid reflections of the pertinent realities of medical care, as well as the values placed by patients on the costs and benefits associated with it.[15] Differences in values would result in different levels of satisfaction, even when other factors in the medical settings evaluated were held constant. Therefore, the value of continuity for different groups of patients can be assessed in terms of its weight in producing change in satisfaction. If this weight is greater in one group than in another, all other things being equal, it can be inferred that it is valued more in the one group than in the other.

Methods

Sample of Children with Disabilities (SCWD)

Families of children with cystic fibrosis, cerebral palsy, myelodysplasia and multiple handicaps were selected from the case loads of four pediatric specialty clinics in two teaching hospitals in Cleveland, Ohio. The majority of children with these conditions are referred to tertiary medical institutions. Since the participating hospitals are two of the three tertiary hospitals in the Cleveland area, their clinics' case loads provide relatively representative samples of area children in these diagnostic categories. The four clinics provide care for all cystic fibrosis patients and the majority of patients with myelodysplasia in the area. Although the proportion of area children with cerebral palsy cared for in these clinics cannot be estimated, they come from a wide range of socioeconomic backgrounds and referral networks. Multiple handicaps comprise a wide variety of congenital disorders, all with physical stigmata, some of which are identifiable as specific syndromes. Although the exact proportion of such children in the community who receive care in these clinics cannot be determined, it may well be the majority.

All families of active patients between three and eighteen years of age in these diagnostic categories who resided in the Cleveland area were asked to participate in the study. From 460 eligible families, 369 (80 per cent) complete and usable interviews were obtained. † The sample includes 77 families of children with cystic fibrosis, 118 with cerebral palsy, 82 with myelodysplasia and 92 with multiple physical handicaps.

Sample of Children in the Community (SCHC)

A three-stage probability sample was designed to represent all area (Cuyahoga County) families with children between three and 18 years of age. The 1977 Cleveland Area Survey (CAS) sampling scheme and lists of blocks and addresses were used for drawing this sample. The CAS classified all blocks (using updated 1970 Census data) into six strata according to: 1) racial composition (less or more than 50 per cent blacks) and 2) average housing value (high, medium, low). Three hundred blocks were randomly selected from these six strata, each stratum providing a fraction proportional to the total population that the stratum encompassed. Five households were randomly selected from each block, yielding 1,500 households to be screened as to whether they contained a child between the ages of three and 18 years and thus were eligible for the study. From 534 eligible families, 459 (85 per cent) complete interviews were obtained.

For the medical care of their children, 75 per cent of the families used private pediatric offices in the community; the remaining families used outpatient departments of local hospitals, a large prepaid medical group practice, city- and county-owned clinics, or a variety of other facilities in the area.

Data and Measurement

Mothers were interviewed in their homes using a structured questionnaire with several self-administered instruments. Trained female interviewers, matched according to respondent race, conducted the interviews. Information was obtained on a wide range of areas, including sociodemographic characteristics, the health status of an Index Child (the disabled child in the SCWD and a randomly selected child between three and 18 years of age in the SCHC), health services used on his or her behalf, and satisfaction with medical care, i.e., the specialty clinics in the SCWD or a variety of medical settings in the SCHC.

---

† In a previous article from this study (Breslau and Mortimer, 1981) and in reports of other studies in which the PSQ was used, items were scored from 1 to 5. Translation of scores by shifting them to a 0 to 4 range was employed to simplify the presentation of the statistical analysis.

Satisfaction with medical care was measured by the Patient Satisfaction Questionnaire (PSQ), developed by Ware et al.[15,16] The PSQ consists of 43 statements that form eight subscales. Respondents were asked to state whether they agreed or disagreed with each statement on a 5-point scale ranging from strongly agree (scoring 4) to strongly disagree (scoring 0). † Scale scores were computed by summing the scores for items contained in each scale, after reversing the codes of replies to unfavorably worded items. Scale scores thus measure approval of, or satisfaction with, the medical care evaluated: the higher the score, the greater the satisfaction.

Continuity was measured by respondents' reports of the extent to which their visits to the clinic (or doctor's office) were attended by the same doctor. The three-item continuity scale (with possible scores of 0 to 12) was used. The three items are descriptive rather than evaluative, and as such do not measure satisfaction with continuity but instead continuity of physician itself.[17] On the basis of a review of previous studies and their own research, Ware et al.[15] concluded that this scale is a valid measure of continuity of physician. Internal consistency reliabilities (Cronbach's alpha) were 0.53 in the SCWD and 0.50 in the SCHC. Three scales—Quality of Care (with seven items and possible scores of 0 to 28), Humaneness of Doctors (with eight items and possible scores of 0 to 32) and General Satisfaction (with four items and possible scores of 0 to 16)—were used as dependent variables in this analysis. Reliabilities (Cronbach's alpha) were, in the SCWD and SCHC, respectively, for Quality, 0.67 and 0.67; for Humaneness, 0.78 and 0.81; and for General Satisfaction, 0.82 and 0.70. (See Appendix for scale items.) These scales measure patient evaluations of the affective and instrumental dimensions of the medical care they received and their general satisfaction with it. Other scales of the PSQ refer to more specific conditions of medical care and are therefore less appropriate indicators of the impact of continuity of physician on patient attitudes.

Three patient characteristics—education, income and race—were used as control variables. Mother's education was measured by years of schooling completed, and income was defined as total family income for 1977. Race was measured as a dummy variable, with a value of 1 assigned to whites and 0 to nonwhites. Patient estimate of waiting time in the doctor's office was also used as a control variable. Previous research has found consistently that patient satisfaction was affected by waiting time,[17,18] and, since long waits often characterize clinic settings in which continuity of physician tends to be low, the effect of continuity on patient satisfaction would be more accurately measured when the influence of waiting time is controlled. A single item from the PSQ "People are usually kept waiting a long time when they are at the doctor's office," measured waiting time in this analysis. Possible scores on this scale were from 0 to 4. In a previous article, data were presented to support the construct validity of this measure.[14]

Analysis

The analysis reported here uses a multiple regression approach designed for the study of interaction.[19] The hypothesis to be tested is that of an interaction effect: the impact of continuity of physician on patient satisfaction is greater in the sample of disabled children (SCWD) than in the randomly selected community sample (SCHC). Such an interaction effect can be tested and estimated by adding, in a hierarchical order, a product term (group membership × continuity) to the regression equation:

$$\bar{y} = a + b_1x_1 + b_2x_2 + b_3x_1x_2,$$

where $\bar{y}$ = satisfaction, x_1 = group membership, x_2 = continuity. When group membership is defined as a dummy variable, with a value of 0 assigned to SCHC and of 1 to SCWD, the specific interaction hypothesis would predict that: 1) b_3 is statistically significant, and the increment in the proportion of variance of $\bar{y}$ accounted for by the interaction term is significant; and 2) the signs of the coefficients b_2 and b_3 are positive, indicating that the continuity slopes in both groups are positive and that the slope is steeper in the SCWD than in the SCHC. If the interaction is significant, the regression coefficient of satisfaction on continuity can be calculated for each group from the overall regression results as follows: the intercept for the overall regression, a, measures the intercept for the regression equation for the SCHC (i.e., the group assigned 0). The intercept of the regression equation for the SCWD (i.e., the group assigned 1) is equal to the intercept for the overall regression, a, plus the coefficient for group membership, b_1. The partial regression coefficient of the interaction term, b_3, measures the difference between the regression coefficients of continuity in the SCWD and in the SCHC, and the partial regression coefficient of continuity in the overall regression, b_2, measures the regression coefficient of continuity in the SCHC. The regression coefficient for continuity in the SCWD is, therefore, equal to b_2 plus b_3. From these separate regressions, one can calculate the point of intersection of the two regression lines and determine whether the interaction is ordinal or disordinal, i.e., whether the rank order of the two groups with respect to satisfaction is constant or whether it is reversed with changes in continuity.

The same analytic procedure can be employed when the model includes control variables:

$$\bar{y} = a + b_1c_2 + b_2c_2 + b_3c_3 + b_4c_4 + b_5x_1 + b_6x_2 + b_7x_1x_2,$$

where $\bar{y}$ = satisfaction, c_1 to c_4 = control variables (i.e., education, income, race, waiting time), x_1 = group membership, x_2 = continuity.

Results

Table 1 presents descriptive statistics for the variables in the full model. The two groups—the sample of disabled children (SCWD) and the sample of children in the community (SCHC)—varied in maternal education, family income and racial composition, as can be seen in Table 1. The SCHC had, on the average, a higher level of maternal education than the SCWD, 12.6 and 12 years of schooling, respectively. In 1977 dollars, the SCHC had higher mean family income than the SCWD, $21,000 and $16,000, respectively. The proportion of non whites in the SCHC was higher than in the SCWD, 31 and 24 per cent, respectively.

Waiting time in the doctor's office or clinic had, on the average, similar scores in the two groups, 2.69 in the SCWD and 2.62 in the SCHC. This finding suggests that patients of hospital clinics did not regard the waiting time in these settings as more burdensome than did patients their waiting time in other settings, commonly believed to be more convenient and more accommodating to patients.

An inspection of Table 1 also reveals that medical care received by disabled children was less continuous, compared with care received by normal children, scoring 6.82 and 7.88, respectively. The difference in mean continuity between the groups, although statistically significant, was relatively small. The range of the observed scores was similar in the two groups, from 0

TABLE 1. Zero-Order Correlations, Means, Standard Deviations and P-Values of Control, Independent and Dependent Variables for SCWD* and SCHC*

	Zero-Order Correlations†								Means and SDs‡		
Variables	1	2	3	4	5	6	7	8	SCWD (n = 369)	SCHC (n = 459)	P-Value§
1. Income	—	0.36	−0.25	0.22	0.08	0.02	0.14	0.10	16,000 (11,449)	21,000 (15,476)	<0.001
2. Education	0.40	—	−0.12	0.10	0.12	0.05	0.08	0.06	12.00 (2.40)	12.60 (2.14)	<0.001
3. Race (% nonwhites)	−0.33	−0.21	—	−0.11	−0.01	0.04	0.00	−0.07	24 —	31 —	<0.001
4. Waiting Time	0.10	0.13	−0.14	—	0.33	0.40	0.48	0.49	2.69 (1.22)	2.62 (1.14)	NS
5. Continuity	0.12	0.18	−0.26	0.26	—	0.40	0.50	0.44	6.82 (2.26)	7.88 (1.97)	<0.001
6. Satisfaction with Quality	−0.00	0.05	0.01	0.30	0.22	—	0.75	0.71	18.61 (3.99)	17.67 (3.82)	<0.001
7. Satisfaction with Humaneness	−0.03	−0.02	0.01	0.33	0.23	0.66	—	0.77	21.68 (5.33)	20.30 (4.66)	<0.001
8. General Satisfaction	−0.02	0.02	−0.05	0.34	0.26	0.64	0.65	—	10.74 (3.26)	10.63 (2.41)	NS

* SCWD: Sample of Children with Disabilities; SCHC: Sample of Children in the Community.

† Top triangle = SCWD; bottom triangle = SCHC.

‡ Standard Deviations shown in parentheses.

§ The p-value for RACE is shown from a chi-square test. All others are based on t-tests of the difference between two sample means.

to 12 in SCWD and from 1 to 12 in the SCHC, indicating that the levels of continuity experienced by patients in the two groups overlapped almost completely.

A comparison of the two groups on the three satisfaction scales, presented in Table 1, indicates that mothers of disabled children cared for in the specialty clinics scored higher than mothers from randomly selected households on Quality of Care and Humaneness of Doctors. The differences were relatively small but statistically significant. On General Satisfaction, the two groups were virtually the same; and the minor difference was not statistically significant.

These findings taken together are somewhat puzzling: of the two groups, the one with the lower mean continuity—SCWD—had, on the average, higher scores of satisfaction with doctors' attitudes and competence. On the surface, they disconfirm the proposition that continuity promotes patient satisfaction. One may speculate that the differences in patient demographic characteristics account for the unexpected results: the community sample includes more families from high socioeconomic strata, who may be less satisfied because they hold higher expectations. It will be shown, however, that the findings at the aggregate level are misleading and that the explanation for them lies elsewhere.

The hypothesis that patient satisfaction was not equally associated with continuity of physician in the two groups was tested. In three separate regression analyses, using the combined sample ($n = 828$), the three patient satisfaction scales were regressed on 1) group membership; 2) continuity; and 3) a product term of group membership × continuity. Group membership was defined as a dummy variable, with a value of 1 assigned to SCWD and 0 to SCHC. Results from the three analyses, presented in Table 2, provide general confirmation for the hypothesis. Ordinary least-squares unstandardized regressions are presented in the table; coefficients whose values exceed twice their standard errors are marked as asterisks as statistically significant. For all three satisfaction scales, the regression coefficients for the interaction of continuity with group membership, b_3, were statistically significant. Table 2 also shows that the coefficients for continuity, b_2, which measure the regression lines in the SCHC, were also statistically significant. From their signs, it can be inferred that in the SCHC continuity was positively related to all three patient satisfaction scales. All three regression coefficients of the interaction term, b_3, were also

TABLE 2. Regressions of Satisfaction Scales on Continuity, Group Membership and Continuity × Group Membership ($n = 828$)
(unstandardized regression coefficients; standard errors in parentheses)

	Independent Variables				
Dependent Variables	Group (b_1)	Continuity (b_2)	Continuity × Group (b_3)	Constant (a)	R^2
Quality	−1.344 (1.296)	0.424* (0.087)	0.276* (0.122)	13.057	0.112
Humaneness	−4.299* (1.603)	0.541* (0.108)	0.636* (0.151)	14.410	0.170
General Satisfaction	−2.581* (0.917)	0.321* (0.062)	0.308* (0.086)	7.137	0.142

* Coefficient exceeds twice its standard error.

positive, indicating that in each area of satisfaction the difference in continuity weight between SCWD and SCHC was one of excess for SCWD. Consequently, it can be deduced that the weight of continuity in producing change in each of the three patient satisfaction areas was greater in the SCWD than in the SCHC. The model explained 11 per cent of the variance in satisfaction with Quality of Care, 17 per cent in satisfaction with Humaneness of Doctors and 14 per cent in General Satisfaction.

To describe the nature of the interactions more clearly, regression equations for each group were calculated from the results of the overall regressions. These are presented in Table 3. In all three areas, patient satisfaction was considerably more strongly dependent on continuity of physician in the sample of disabled children (SCWD) than in the community sample (SCHC). In Humaneness of Doctors, the excess impact of continuity in the SCWD was the greatest: the coefficient for continuity in this group was 1.177, compared with 0.541 in the SCHC, indicating that the change in satisfaction with Humaneness associated with a unit change in continuity in the SCWD was more than twice as large as a corresponding change in the SCHC. In satisfaction with Quality of Care, the excess impact of continuity in the SCWD was the smallest, but even there it was one and two-thirds times greater than in the SCHC, 0.700 and 0.424, respectively. In General Satisfaction, the impact of continuity in the SCWD was almost twice as great as in the SCHC. In Table 3 appear also the values of the intercepts of the regressions in the SCWD and in the SCHC. In all three areas of patient satisfaction, the intercepts were lower in the SCWD than in the SCHC, indicating that at zero continuity, families of disabled children were less satisfied than were families of the community sample.

The efficacy of continuity in predicting satisfaction was markedly greater in the SCWD than in the SCHC, according to the correlations of continuity with the three satisfaction scales (Table 1). The values of continuity at the point of intersection were calculated from the intercepts and the regression coefficients for continuity in Table 3. The value of continuity at the point of intersection of the two regression lines was well within the range of interest for each of the three dependent variables in this study. Thus, for each area of patient satisfaction, the interaction was disordinal: at low levels of continuity, those in SCHC were more satisfied, but at high levels of continuity, those in SCWD were more satisfied. A clear picture of these results can be seen by plotting the regression lines for the two groups, as was done in Figure 1.

The hypothesis that the impact of continuity on satisfaction was stronger in the sample of disabled children (SCWD) than

TABLE 3. Estimates of Separate Regression Equations of Three Satisfaction Scales on Continuity in SCHC* and SCWD* (unstandardized regression coefficients)†

Dependent Variables	SCHC		SCWD		Continuity at Point of Intersection
	Constant	Continuity	Constant	Continuity	
Quality	13.057	0.424	11.723	0.700	4.833
Humaneness	14.410	0.541	10.111	1.177	6.759
General Satisfaction	7.137	0.321	4.556	0.629	8.399

* SCHC: Sample of Children in the Community; SCWD: Sample of Children with Disabilities.

† All differences between corresponding regression coefficients are statistically significant ($p < 0.05$).

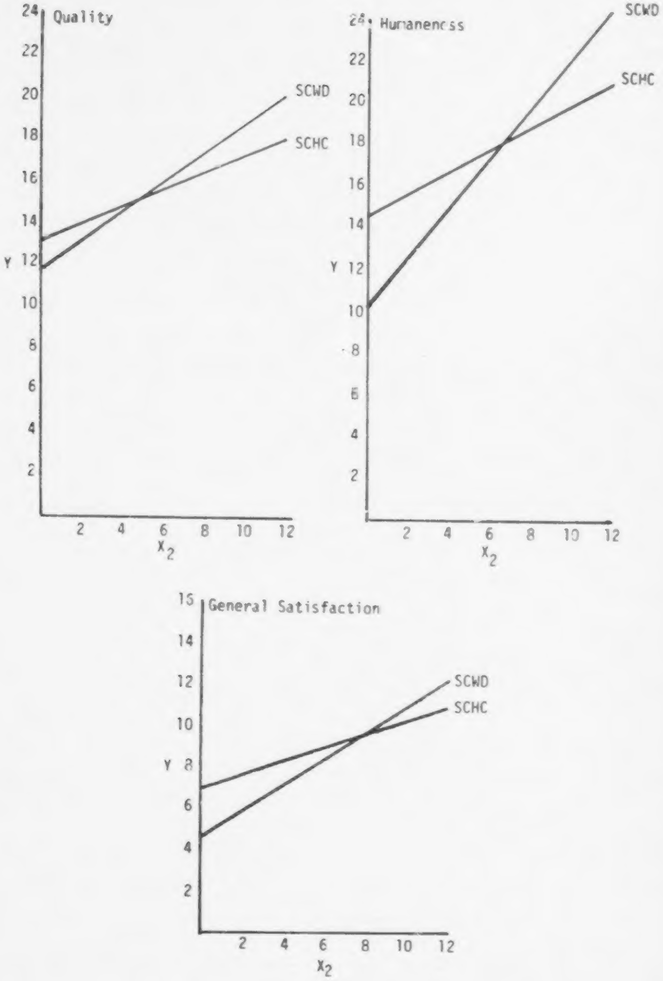


FIG. 1. Regressions of satisfaction (Y) on continuity (X_2) for SCWD and SCHC.

in the community sample (SCHC) was tested also net of the effects of maternal education, family income, race and waiting time upon arrival. In three regression analyses using the combined sample (n = 828), satisfaction was regressed on mothers' education, family income, race, waiting time, group membership, continuity and a product term of group membership × continuity. The results of these analyses are presented in Table 4. As can be seen from this table, none of the sociodemographic variables had statistically significant regression coefficients in any of the three equations. In contrast, waiting time had statistically significant coefficients in all three regression equations. Their positive signs indicate that shorter waiting time produced an increase in patient satisfaction with Quality of Care and Humaneness of Doctors, as well as in General Satisfaction with medical care. The

TABLE 5. Estimates of Separate Regression Equations of Three Satisfaction Scales on Continuity in SCHC* and SCWD,* Controlling for Education, Income, Race and Waiting (unstandardized regression coefficients)†

Dependent Variables	SCHC		SCWD		Continuity at Point of Intersection
	Constant	Continuity	Constant	Continuity	
Quality	11.652	0.307	10.487	0.539	5.021
Humaneness	13.042	0.371	8.976	0.950	7.022
General Satisfaction	6.840	0.197	4.216	0.485	9.111

* SCHC: Sample of Children in the Community; SCWD: Sample of Children with Disabilities.

† All differences between corresponding coefficients are statistically significant ($p < 0.05$).

Discussion

The results of this analysis confirm the hypothesis that the impact of continuity on satisfaction with the medical care provided to their children was greater in a sample of families of disabled children than in a sample of randomly selected families. In regard to key dimensions of medical care, continuity of physician influenced favorably the evaluations of families in both groups: as continuity rose, so did their approval of the quality of care, the humaneness of doctors and the overall level of medical care. The greater impact that continuity had on the attitudes of families of disabled children, compared with those of randomly selected families, meant that when medical care was discontinuous, families of disabled children were less satisfied than the randomly selected families, but when continuity was at a high level, their satisfaction was greater. The results thus indicate that families with disabled children, who attend specialty clinics, place greater value on seeing the same doctor from visit to visit than do families whose children are free of disabilities.

The analysis focused on the hypothesis that the impact of continuity on patient satisfaction is greater in families with disabled children than in the general population; it did not attempt to specify the factors that might explain this difference. Clearly, compared with children in the general population, disabled children have more severe health problems, which bring them to doctors' offices more frequently. Future research should attempt to estimate the extent to which these and other factors account for the relationship between continuity and satisfaction in these families. With respect to one such factor, namely differences in severity among the disabled children, a previously reported analysis from this study failed to find significant effects on satisfaction.[14] It should be kept in mind, however, that *all* the disabled children in the study had life-threatening or severely handicapping conditions, in spite of wide variation in their physical limitations at the time of the study.

It may be that, due to the distress experienced by families of disabled children, medical care serves for them an important function beyond the technical procedures it brings to bear on the diagnosis and treatment of the illness. Parsons[20] suggested that patients attribute to their physicians significance as persons that may not be directly relevant to their technical functions. Although this becomes of key importance in psychotherapy (where it is referred to by the term transference), it is always a factor in doctor-patient relationships. Its importance, according to Parsons, is proportional to the duration of the relationship and, even more so, to the emotional importance of the health problem and, hence, the relationship to the physician. Other writers have emphasized the importance of the severity of the illness as an influence on patients' attitudes toward their physicians.[21,22] Specifically, it was suggested that the affective dimension in a doctor–patient relationship looms larger as the patient's illness arouses higher levels of anxiety. Similar arguments were advanced recently by Miller.[23] He proposed that in severe childhood illness, family and child look to the physician to fulfill an important psychologic function in the child's development and the adjustment of the family to the disease. A frequent consequence of severe disability in a child is social isolation, as the family withdraws from extradomestic activities and as neighbors, friends and sometimes kin avoid contacts with the family. In this restricted social context, the physician fulfills important functions of providing information and emotional support and strengthening positive patterns of family relationships and child-rearing practices.

The excess impact of continuity of physician on the evaluations that families of disabled children made of their doctors and medical care may reflect the greater psychologic needs these families bring with them to their doctors, compared with families of normal children. A doctor–patient relationship conducive to the expression of psychologic needs and their resolution is more critical for these patients than for patients with self-limiting illnesses. To the extent that continuity of physician promotes such doctor–patient relationships, it would indeed be more highly valued by families of disabled children than by families with children free of disabilities.

The findings of this study have important implications for the organization of medical care in specialty clinics. The entire sample of disabled children was drawn from the case loads of four specialty clinics, representing a variety of types of specialty medical care. Two of the clinics provide to their patients a full range of medical services, those connected with the disability as well as all immunizations and care for intercurrent childhood illness. The third clinic focuses only on the disability, its neurologic and orthopedic diagnosis and management, whereas in the fourth, although staff neither administer immunizations nor provide care for intercurrent illness, the concern with the disability includes its developmental consequences. The clinics vary also in the extent to which the care they provide is continuous. In one clinic, each physician–specialist has a distinct case load and is available to patients on any day of the week. Continuity of physician is at a high level at this clinic, far higher than at the other clinics in which similar arrangements to assure continuity have not been instituted. In two of these clinics, arrangements for continuity would have required intricate coordination, since they involved large teams of specialists from many fields. These latter clinics are examples of a growing number of subspecialty clinics providing technically advanced medical care to children with chronic illness or disabilities. The findings that continuity of physician had a far greater impact on the attitudes of patients of specialty clinics than on patients who receive care in primary care settings calls attention to the organizational conditions that characterize these clinics. Continuity of physician in these clinics is not a natural consequence of their organizational structure, work schedules and staffing. To secure continuity, new work arrangements are required. The technical benefits to be derived from the involvement of an array of specialists must be weighed against the costs of discontinuity. In the eyes of the disabled patients and their families, continuous care is not only a demonstration of doctors' concern, but also of the quality of care they provide.

References

1. Alpert JJ, Charney E. The Education of Physicians for Primary Care. Rockville, MD: Department of Health, Education and Welfare, 1973. (Public Health Service Publication No. (HRA) 74-3113.)

2. Hansen MF. Continuity of care in family practice. Part 3: Measurement and evaluation of continuity of care. J Fam Pract 1975;2:439.

3. Steinwachs DM. Measuring provider continuity in ambulatory care: an assessment of alternative approaches. Med Care 1979;17:551.

4. Alpert JJ, Heagarty MC, Robertson LS, et al. Effective use of comprehensive pediatric care. Am J Dis Child 1968;116:529.

5. Alpert JJ, Kosa J, Haggerty RJ, et al. Attitudes and satisfaction of low-income families receiving comprehensive pediatric care. Am J Public Health 1970;60:499.

6. Becker MH, Drachman RH, Kirscht JP. Continuity of pediatrician: new support for an old shibboleth. J Pediatr 1974;84:599.

7. Breslau N, Haug MR. Service delivery structure and continuity of care: a case study of a pediatric practice in process of reorganization. J Health Soc Behav 1976;17:339.

8. Heagarty MC, Robertson LS, Kosa J, Alpert JJ. Some comparative costs in comprehensive versus fragmented pediatric care. Pediatrics 1970;46:596.

9. Langston JH, et al. Study to Evaluate the OEO Neighborhood Health Center Program at Selected Centers. Rockville, MD: Geomet, Inc., 1972.

10. Sussman MB, Caplan EK, Haug MR, Stern MR. The walking patient: a study in outpatient care. Cleveland: Press of Western Reserve, 1967.

11. McDermott W. General medical care: identification and analysis of alternative approaches. Johns Hopkins Med J 1974;135:292.

12. Aiken LH, Lewis CE, Craig J, et al. The contribution of specialists to the delivery of primary care. N Engl J Med 1979;300:1363.

13. Starfield B. Continuous confusion? Am J Public Health 1980;70:117.

14. Breslau N, Mortimer EA Jr. Seeing the same doctor—determinants of satisfaction of "specialty" care for disabled children. Med Care, 1981;19:741.

15. Ware JE Jr, Snyder MK, Wright WR. Development and Validation of Scales to Measure Patient Satisfaction with Health Care Services: Volume I of a Final Report. Prepared for the National Center for Health Services Research, Health Resources Administration, U.S. Department of Health, Education, and Welfare, 1976.

16. Ware JE Jr, Snyder MK. Dimensions of patient attitudes regarding doctors and medical care services. Med Care 1975;13:669.

17. Aday LA, Andersen R, Fleming GV. Health care in the United States: equitable for whom? Beverly Hills: Sage Publications Inc, 1980.

18. Tessler R, Mechanic D. Consumer satisfaction with prepaid group practices: a comparative study. J Health Soc Behav 1975;16:95.

19. Kerlinger FN, Pedhazur EJ. Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston, Inc, 1973.

20. Parsons T. The social system. New York: Free Press, 1951.

21. Ben-Sira Z. The function of the professional's affective behavior in client satisfaction: a revised approach to social interaction theory. J Health Soc Behav 1976;17:3.

22. Ben-Sira Z. Affective and instrumental components in the physician–patient relationship: an additional dimension of interaction theory. J Health Soc Behav 1980;21:170.

23. Miller D. CF: Its impact on the individual, the family, and the caregiver. Proceedings of the Cystic Fibrosis Foundation GAP Conference, Colorado Springs, Colorado, July 1980, Volume 5, No 1, 1981.

Appendix

The Three Subscales From the Patient Satisfaction Questionnaire Used as Dependent Variables

Quality of Care

Doctors aren't as thorough as they should be.

Most people are encouraged to get a yearly exam when they go for medical care.

Sometimes doctors take unnecessary risks in treating their patients.

Doctors are very careful to check everything when examining their patients.

Doctors ask what foods patients eat and explain why certain foods are best.

The medical problems my child had in the past are ignored when we seek medical care for a new medical problem.

Doctors don't advise patients about ways to avoid illness or injury.

Humaneness of Doctors

Doctors always do their best to keep the patient from worrying.

Doctors always treat their patients with respect.

Sometimes doctors make the patient feel foolish.

Doctors always avoid unnecessary patient expense.

Doctors cause people to worry a lot because they don't explain medical problems to the patients.

Doctors respect their patients' feelings.

Doctors never recommend surgery (an operation) unless there is no other way to solve the problem.

Doctors hardly ever explain the patient's medical problems to him/her.

General Satisfaction–Dissatisfaction

I'm very satisfied with the medical care my child receives.

The care my child has received from doctors in the last few years is just about perfect.

Most people receive medical care that could be better.

There are things about the medical care my child receives that could be better.

Continuity of Care

My child hardly ever sees the same doctor when he/she goes for medical care.

If more than one family member needs medical care, we have to go to different doctors.

My child sees the same doctor just about every time he/she goes for medical care.

MEDICAL CARE
April 1982, Vol. XX, No. 4

Distress Syndromes, Illness Behavior, Access to Care and Medical Utilization in a Defined Population

DAVID MECHANIC, PH.D.,* PAUL D. CLEARY, PH.D.,†
AND JAMES R. GREENLEY, PH.D.,‡

This article examines the use of general medical services in a representative sample from a defined geographic area and in a sample of persons seeking psychiatric care from the same area. Psychiatric patients made 100 per cent more general medical care visits in the retrospective period and 83 per cent more in the prospective period than persons who did not seek mental health care. The analysis focuses on the determinants in general medical care use between those who sought mental health care and those who did not. The first hypothesis is that physical symptoms and dysfunction concomitant with psychologic disorder explain the difference. The second argues that the association is a product of help-seeking orientations and illness behavior. The third focuses on variations due to differences in access. The first two types of factors are the most important. Using sex, physical symptoms and illness behavior measures, we explain 50 per cent of the differences in retrospective utilization and 40 per cent of the differences in prospective data.

WITH GROWING INTEREST in the relationship between general care and mental health,[1] attention has been refocused on the associations between psychiatric and physical morbidity and related help-seeking. From a policy perspective, there is interest in various studies suggesting that the provision of psychologic services reduces demands on the medical care system, and on the use of costly and possibly dangerous interventions.[2-3] Since the general medical sector is the major source of care for persons with psychologic problems and since such persons use a disproportionate amount of service, the indications that psychologic management of such persons reduces utilization are of major policy interest.

The literature on the relationship between psychologic morbidity and general medical utilization, however, is more uncertain than the results of such studies might indicate. All existing studies have major selection biases affecting which distressed persons receive psychologic management, making it difficult to interpret the significance of the utilization reduction. There is also a sizable literature on medical

---

* University Professor and Dean, Faculty of Arts and Sciences, Rutgers University, New Brunswick, New Jersey.

† Associate Research Professor of Social Work, Rutgers University, New Brunswick, New Jersey.

‡ Associate Professor, Departments of Psychiatry and Sociology, University of Wisconsin, Madison.

Supported in part by grants from the Center for Epidemiological Studies, NIMH, and from the Robert Wood Johnson Foundation. An earlier version of this paper was presented at the symposium, What is a Case?, Institute of Psychiatry, London, May 1980.

Address for reprints: David Mechanic, University Professor and Dean, Faculty of Arts and Sciences, Rutgers University, 77 Hamilton Street, New Brunswick, NJ 08903.

0025-7079/82/0400/0361/$01.10 © J. B. Lippincott Co.

utilization indicating that patients receiving psychiatric treatment, as compared with those who do not, have higher levels of medical care utilization.[4–5] Given the importance of correct interpretation for policy decisions, as well as for understanding, we cannot simply dismiss the contradictions apparent in these two areas of literature.

The research literature referred to above supports the following general conclusions: 1) there is an association between reports of psychologic and physical distress, and persons with psychologic distress feel sicker and subjectively rate their physical health as less adequate[6]; 2) physical symptoms and psychologic distress both independently contribute to seeking general medical care[7]; 3) patients in psychiatric treatment use medical care of all kinds more frequently than those not receiving such care[4]; and 4) in some settings the provision of psychologic management is related to a reduction of medical utilization.[2]

These findings are amenable to alternative interpretations. The purpose of the present study was to examine in a defined population how patient status, distress, access to care and other variables are related to retrospective and prospective medical utilization (excluding psychiatric care).

The Research Setting

In 1977, we initiated a study of persons in north central Wisconsin who were served predominantly by a large multispecialty group practice and satellite clinics. Analyses presented here are based on two samples drawn from a defined area with a population of approximately 50,000§: 1) a representative sample of 1,026 persons 18 years of age or over in the population; and 2) a consecutive sample of 91 patients coming to the psychiatry department of the clinic for mental health services.‖ Eligibility for the second sample—referred to as the psychiatry sample—was based on four criteria: the patient had to 1) reside in the population area from which we drew our representative community sample; 2) be over 18 years of age; 3) have no record of consulting the psychiatric clinic in the previous 30 days; and 4) have a scheduled appointment of more than 15 minutes. The latter two criteria were used to define new episodes,# with the 15-minute criterion used to eliminate routine medication checks. We also excluded patients coming directly to obesity, smoking and alcoholism clinics administered by the psychiatric department. In both samples, respondents were interviewed in their homes and asked to complete a questionnaire and return it by mail following the interview. These data were then merged with retrospective and prospective data from respondents' medical and pharmacy records.

Measures of Psychologic Distress

Our data include a large number of scales of psychologic distress intended to

---

§ Because we did not use a simple random sample, in some cases our calculations may slightly overestimate the p-values of significance tests.

‖ Psychiatric patients were concerned with confidentiality, and twenty-eight patients asked not to be included in the study. Because such a high refusal rate could bias our results, we compared our sample with those who did not participate in the study to see how they differed on a variety of dimensions. There were no significant differences between our sample and those who refused to participate in presenting diagnosis, number of visits to the psychiatry department over a two-year period, mean age, marital status or sex. These two groups did differ with respect to past psychiatric history. Eight per cent of our sample had been hospitalized prior to the study for a psychiatric problem in contrast to 43 per cent of those who refused to participate. It appears that the patients we lost had either more severe or more chronic conditions than those included in the sample.

# Twenty-three of the patients in our psychiatric sample had indications in their charts that they had sought psychiatric care at some time in the past. We examined how these patients differed from patients without prior episodes on 50 variables, and the only statistically significant difference was that those with a prior episode had slightly less education.

measure varying types of psychopathology and expression patterns. In fact, these tend to be highly intercorrelated, and, thus, we focus on only six summary measures. The first, based on measures developed by Dohrenwend, which he calls *demoralization*,[8] is a summary score based on eight symptom scales most highly intercorrelated in Dohrenwend's PERI (Psychiatric Epidemiology Research Interview) questionnaire. The eight scales are self-esteem, hopelessness, dread, confused thinking, sadness, anxiety, psychophysiologic symptoms and perceived physical health. Scale construction followed Dohrenwend's procedures with slight modifications.

The second measure was a 22-item Langner scale.[9] The version used in this study was scored such that it had a range of 22 to 110[¶] and in the representative sample had an internal consistency (coefficient Alpha) of 0.87. The third measure was a Distress Summary Score (I), based on 17 scales, most adapted from Dohrenwend, measuring drinking problems, psychophysiologic symptoms, problems that are possibly psychogenetic, neurotic problems, obsessive symptoms, guilt, suicide-scale, enervation, anxiety, sadness, hopelessness and fears. Each scale was standardized, and these standardized scale scores were summated. A second Distress Summary Score (II) was constructed in a similar fashion, eliminating items from five scales (psychophysiologic, possible psychogenic, sleep problems, anxiety and enervation). The purpose of this fourth scale was to have a general psychologic distress indicator unconfounded by physical or psychophysiologic content. Two additional scales included in the analysis were the Eysenck Neuroticism Scale (coefficient Alpha = 0.55) and a Depression Scale composed of four previously mentioned scales (guilt, sadness, hopelessness and suicide). These six scales achieve adequate levels or reliability,[**] and have been found to be related to levels of functioning and illness behavior.

We examined the intercorrelations among these five measures for the cross-sectional and psychiatric samples (Table 1). These measures share many of the same items, and thus high intercorrelations are expected. With the exception of the

TABLE 1. Correlations Between Distress Indicators
Psychiatric Sample (N = 91) Above Diagonal
Area Probability Sample (N = 1,026) Below Diagonal, Pairwise Deletion Used

	1	2	3	4	5	6
1. Demoralization	—	.93	.94	.86	.60	.81
2. Langner	.91	—	.89	.76	.63	.66
3. Distress summary I	.93	.90	—	.96	.60	.86
4. Distress summary II (psychologic)	.87	.80	.97	—	.55	.92
5. Eysenck	.58	.59	.56	.51	—	.45
6. Depression scale	.88	.78	.89	.90	.52	—

[¶] For each item in the Langner scale, respondents were asked whether they had experienced that symptom very often, fairly often, sometimes, almost never or never in the prior three months. The responses to these items, which were scored as 5, 4, 3, 2 or 1, were then summed to form the Langner score.

[**] The internal consistencies (coefficient Alpha) for the components of the various summary measures are: self-esteem (0.72), hopelessness (0.78), dread/fear (0.43), confused thinking (0.79), sadness (0.81), anxiety (0.82), psychophysiologic symptoms (0.52), perceived physical health (0.45), drinking problems (0.77), manic behavior (0.56), delusions (0.52), explosiveness (0.72), sleep problems (0.63), possible psychogenic problems (0.60), neurotic problems (0.69), obsessive symptoms (0.70), guilt (0.75), suicide-scale (0.73), enervation (0.75).

TABLE 4. Regressions of Satisfaction Scales on Group, Continuity and Group × Continuity, Controlling for Education, Income, Race and Waiting (n = 828) (unstandardized regression coefficients: standard errors in parentheses)

	Independent Variables								
Dependent Variables	Education (b_1)	Income (b_2)	Race (b_3)	Waiting (b_4)	Group (b_5)	Continuity (b_6)	Continuity × Group (b_7)	Constant (a)	R^2
Quality	0.016 (0.060)	−0.015 (0.010)	0.246 (0.191)	0.947* (0.111)	−1.165 (1.248)	0.307* (0.087)	0.232* (0.117)	11.652	0.187
Humaneness	−0.063 (0.073)	−0.004 (0.012)	0.394 (0.232)	1.367* (0.135)	−4.066* (1.518)	0.371* (0.106)	0.579* (0.143)	13.042	0.266
General Satisfaction	−0.020 (0.041)	−0.007 (0.007)	0.093 (0.136)	0.833* (0.077)	−2.624* (0.863)	0.197* (0.060)	0.288* (0.081)	6.840	0.249

* Coefficient exceeds twice its standard error.

addition of the three sociodemographic factors and Waiting to the model increased the predicted variance in satisfaction with Quality from 11 to 19 per cent, with Humaneness from 17 to 27 per cent and in General Satisfaction from 14 to 25 per cent (compare columns headed R^2 in Tables 2 and 4). Since the combined contribution of education, income and race was less than 1 per cent in each of the three regressions, the additions in explained variance were due almost entirely to waiting time. In spite of the strength of waiting time as a predictor of satisfaction, controlling for its effects did not alter the general results described above. Regression coefficients for the interaction term, b_7 in this model, were all statistically significant. Those for continuity in the SCHC, b_6 in this model, remained also statistically significant. Table 5 presents the regressions for continuity in the two groups, calculated from the overall regressions in Table 4, in which education, income, race and waiting time were controlled. The effects of controlling these variables on the strength of the relationship between continuity and patient satisfaction can be assessed by comparing the regression coefficients in Table 5 with those in Table 3. All the regression coefficients of continuity were somewhat lower when waiting time and the three sociodemographic factors were held fixed. However, the initial results in regard to the relative magnitudes of the impact of continuity in the two groups remained unchanged. The weight of continuity in producing satisfaction with Humaneness in the SCWD was more than two and a half times as large as it was in the SCHC. In General Satisfaction, the ratio of the coefficients for continuity in the two samples was also greater than 2 to 1. In satisfaction with Quality the ratio was 1.7 to 1. There was also little change in the relationship between corresponding intercepts and in the value of continuity at the point of intersection of corresponding regression lines.

Eysenck Neuroticism score, which is correlated 0.5 to 0.6 with the other measures, the remaining measures are very highly correlated in both the cross-sectional (0.8 to 0.9) and psychiatric (0.6 to 0.9) samples suggesting that it probably does not matter a great deal which is used. Eliminating anxiety, psychophysiologic concerns and physical symptom items from the distress summary score had little effect on how this score was correlated with other measures, and the two distress summary scores were correlated 0.97 in the sample of the general population and 0.96 in the psychiatric sample.

The samples were divided into the following groups: I—patients who came to the clinic for psychiatric care (the psychiatric sample); II—respondents in the cross-sectional sample whose medical records revealed a visit to a psychiatrist during the two-year interval in which we reviewed charts; III—those in the cross-sectional sample who had no record of psychiatric visits but who reported visiting a psychiatrist, psychologist, social worker or mental health center for a personal problem in the past year in the interview; and IV—those in the cross-sectional sample with no indication of mental health care. The treated samples had substantially higher scores on each of the six distress indicators, with the psychiatric sample scoring highest on each of the indicators (Table 2).[††]

TABLE 2. Mean Scores of Four Groups on Indicators of Distress* (Standard Deviations in Parentheses)

	Cross-Sectional Sample of Population (Excluding Users of Mental Health Services)	Cross-Sectional Respondents with Psychiatric Record	Cross-Sectional Respondents Who Reported Using Mental Health Services	Psychiatric Sample
	(N = 842)	(N = 42)	(N = 28)	(N = 91)
1. Langner score	81	88	88	95
	(9.6)	(11.1)	(11.1)	(13.4)
2. Depression scale	58	64	67	75
	(8.6)	(9.7)	(15.8)	(14.0)
3. Distress summary score I	−0.41	7.8	8.7	14.9
	(10.3)	(12.2)	(14.7)	(13.7)
4. Distress summary score II	−0.24	4.9	5.9	10.1
	(7.5)	(8.3)	(11.7)	(9.6)
5. Demoralization score mean	5.2	7.8	7.8	10.6
	(3.4)	(4.0)	(4.4)	(5.1)
6. Eysenck neuroticism score mean	1.9	2.7	2.8	3.1
	(1.4)	(1.4)	(1.8)	(1.6)

* For all scales, higher scores correspond to higher symptom levels.

[††] Utilization data, of course, were obtained only for respondents who gave legal consent for such access (89 per cent of the cross-sectional sample). To simplify comparisons, all subsequent analyses are based only on the respondents for whom we have medical record data. Those who did not consent to have their records reviewed were compared with those who did and were found to be slightly older, less educated and with lower incomes than those who gave consent. In addition, they were lower on some of the measures of psychologic distress; were less likely to be involved in clubs and social activities; and were less likely to respond that they trusted their physician, that they felt comfortable in discussing emotional problems with him or that they had friends who had used a psychiatrist. In general, the differences were relatively small and are not likely to greatly influence inferences made from the data because of the high response rate.

Patient Status and Utilization

Medical records were examined for every medical provider named by respondents during the interviews, and we also examined the records of virtually all medical providers in the geographic area to ascertain whether our sample had used services without reporting such utilization. Patients who sought psychiatric care had higher levels of general medical utilization for both the retrospective and prospective periods (Table 3). For example, while nonusers of mental health services had visit rates of 2.6 and 2.4 medical visits for the retrospective and prospective periods, the rates for the psychiatric sample were 5.2 and 4.4, respectively. These differences were only slightly altered when adjusted for differences in age, sex and education among the samples.‡‡

Unlike that in much of the existing literature, the definition of patient status in our study is independent of medical utilization and precedes the measurement of utilization in the case of the prospective data. While the results here are consistent with other studies, the reasons for these findings remain unclear. There are at least three major competing explanations: that medical utilization is a result of physical symptomatology concomitant with psychologic symptomatology; that use of both mental health and general medical services reflect generalized tendencies to seek help or related illness behavior; and that rates of high use of both types of service reflect overall access to both as indicated by method of payment for service and the accessibility of care.

As the first step in exploring alternative explanations, we compared the psychiatric, the cross-sectional respondents receiving mental health care (i.e., reported using mental health services and/or had a psychiatric visit in record), and the cross-sectional sample members who did not receive mental health care with respect to

TABLE 3. Mean Scores on Retrospective and Prospective Medical Care Utilization (Excluding Mental Health) (Standard Deviations in Parentheses)

	Cross-Sectional Sample of Population (Excluding Users of Mental Health Services)	Cross-Sectional Respondents with Psychiatric Record	Cross-Sectional Respondents Who Reported Using Mental Health Services	Psychiatric Sample
	(N = 842)	(N = 42)	(N = 28)	(N = 91)
Retrospective				
Medical utilization mean visits for year prior to interview	2.6	3.3	4.0	5.2
(Standard deviation)	(3.5)	(3.9)	(3.5)	(5.1)
Per cent 3 or more medical visits	35	48	61	57
Prospective				
Medical utilization mean visits for year following interview	2.4	3.8	3.8	4.4
(Standard deviation)	(3.6)	(4.0)	(3.1)	(4.9)
Per cent 3 or more medical visits	31	55	57	58

---

‡‡ Although our analyses focuses on use of outpatient services, people in the psychiatric sample were also likely to be hospitalized. Nonusers of mental health services had an average of 0.10 and 0.11 hospital admissions for the retrospective and prospective periods, while the rates for the psychiatric sample were 0.23 and 0.15, respectively.

five measures of physical functioning, five measures of help-seeking and six measures of access to medical care (Table 4). The physical measures included a two-item subjective assessment of physical condition, a scale of six psychophysiologic symptoms, a scale of nine physical symptoms that are possibly psychogenic, reported number of chronic health conditions, and a measure of bed and hospital disability.§§ Help-seeking measures included three hypothetical scales of tendency to consult physicians (in the case of physical symptoms, in the case of psychologic symptoms and total combined scores), the patient's report of the importance of physical checkups and the patient's acknowledgment that it is appropriate to talk to doctors about personal problems. Access-to-care measures included whether the patient was a member of the prepaid plan, whether the patient reported having a personal doctor, whether the patient could name a specific doctor, convenience in getting to the doctor, difficulty getting an appointment and the reported

TABLE 4. Mean Scores on Measures Reflecting Alternative Hypotheses Accounting for Utilization Differences

	Cross-Sectional Sample		
	Non-Mental Health Users	Mental Health Users	Psychiatric Sample
	(N = 842)	(N = 70)	(N = 91)
Hypothesis 1—Physical Morbidity			
Perception of Physical Condition	5.1	5.5	5.6*
Psychophysiologic symptoms	20.5	21.6	22.8*
Possible psychogenic symptoms	29.1	30.6	31.3*
Number of chronic health conditions	2.6	3.4	3.7*
Bed and hospital disability	0.5	0.7	0.9*
Hypothesis 2—Help-Seeking Tendencies			
Tendency to consult doctor–total	4.3	5.1	4.2*
Tendency to consult doctor–psychologic	0.4	0.5	0.5*
Tendency to consult doctor–physical	4.0	4.6	3.7*
Thinks it important to have checkups	6.1	6.5	7.0*
Thinks it appropriate to talk to doctor about personal problems	2.5	2.8	3.0*
Hypothesis 3—Access to Care			
Proportion naming an individual doctor	0.6	0.7	0.6
Proportion in prepaid group	0.4	0.5	0.7
Has no personal doctor for overall health needs	1.5	1.3	1.4
Inconvenient to get to doctor	1.2	1.2	1.3
Days wait for appointment	23.0	22.3	17.8
Difficulty getting appointment	1.4	1.2	1.4

* One-way analysis of variance among the three groups significant at the 0.05 level.

---

§§ The psychophysiologic symptoms asked about were: poor appetite, constipation, headaches, cold sweats, trembling hands and acid stomach. The possible psychogenic symptoms were: dizziness, fainting spells, heart beating hard, chest pains, shortness of breath, high blood pressure, asthma, diarrhea and weight change without being on a diet. The chronic conditions asked about were: asthma, chronic bronchitis, sinus trouble, rheumatic fever, hardening of the arteries, high blood pressure, heart trouble, stroke, varicose veins, hemorrhoids or piles, hay fever, gall bladder or liver trouble, stomach ulcer, stomach problems, kidney trouble, arthritis, diabetes, thyroid trouble or goiter, allergy, epilepsy, cancer, skin trouble, hernia or rupture, prostate or urinary problems, tremors, paralysis and chronic back problems.

number of days to get an appointment. Each of the physical morbidity and help-seeking measures significantly differentiated the samples, with the psychiatric sample being highest on almost all measures and the cross-sectional respondents who received no mental health care being lowest. The only exception to this pattern was that reported mental health services users had the highest tendency to consult a physician for physical reasons and scored highest on the combined scale of tendency to consult physicians. There were no statistically significant differences among the samples in access to care.

Next, we examined the extent to which age, education and sex differences across the groups explained the utilization patterns. Only sex was a significant predictor of use, with women using more services. The unadjusted differences in medical utilization for one year between the psychiatric sample and the cross-sectional sample members who did not use mental health services was 2.6 visits in the retrospective period and two visits in the prospective period. In other words, the excess of medical visits in the psychiatric sample compared with the cross-sectional non-mental health help seekers was 100 per cent for retrospective data and 83 per cent for prospective data. Adjusting for sex differences among samples reduces the medical utilization differences to 2.3 visits for retrospective data and 1.8 visits for prospective data (Tables 5 and 6).[‖]

[‖] All adjusted means were calculated using analysis of the covariance on the combined, unweighted groups. This procedure assumes that the relationship between the covariates and the dependent variables are the same in each group, an assumption difficult to test with such small groups.

TABLE 5. Medical Utilization Among Groups Controlling for Sex, Illness Behavior and Physical Symptoms

	Cross-Sectional Sample		
	Non-Mental Health Users	Mental Health Users	Psychiatric Sample
	(N = 842)	(N = 70)	(N = 91)
Unadjusted medical care utilization			
Retrospective	2.6	3.6	5.2
Prospective	2.4	3.8	4.4
Medical care utilization adjusted for sex			
Retrospective	2.6	3.5	4.9
Prospective	2.5	3.7	4.3
Medical care utilization adjusted for five physical symptom measures			
Retrospective	2.7	3.2	4.4
Prospective	2.5	3.5	4.0
Medical care utilization adjusted for illness behavior measures			
Retrospective	2.8	3.4	4.6
Prospective	2.6	3.6	4.3
Medical care utilization adjusted for sex and significant physical symptoms* and illness behavior measures			
Retrospective	2.9	3.3	4.2
Prospective	2.7	3.3	3.9

* In retrospective analysis, scale of "symptoms possibly psychogenic" was used; in prospective analysis, perceived physical status was used.

TABLE 6. Differences in Mean Medical Care Utilization Between Cross-Sectional Sample Who Are Nonusers of Mental Health Services and Psychiatric Sample

	Retrospective	Prospective
Unadjusted difference	2.6	2.0
Adjusted by sex	2.3	1.8
Adjusted by physical measures	1.7	1.5
Adjusted by illness behavior measures	1.8	1.7
Adjusted by sex, physical measures and illness behavior	1.3	1.2

We suggested three hypotheses that may account for differences in utilization among these samples: the first explanation accounts for the differences on the basis of magnitude of physical symptoms accompanying psychologic disorder, the second on the basis of differential help-seeking tendencies and the third on the basis of access. Because there were no significant differences among the groups in access, we focused on the first two hypotheses. Thus, we separately adjusted medical utilization rates among the samples by the five physical and psychophysiologic illness scales that differentiated these groups and by the four illness behavior measures on which the groups varied. Controlling for the physical symptom measures reduces the difference of 2.6 visits for retrospective data and two visits for prospective data to 1.7 and 1.5, reductions of 35 per cent and 25 per cent, respectively. Controlling for illness behavior measures reduces the differences to 1.8 and 1.7 visits, reductions of 31 per cent and 15 per cent, respectively. Next, we constructed regression equations to correct utilization for the best combination of significant predictors. In the case of retrospective utilization, four variables—sex, a scale of psysiologic symptoms that could be psychogenic, a measure of bed disability–hospital days, and a measure of the respondent's belief in having regular checkups even when feeling well—reduced group differences for retrospective data to 1.3 visits, a reduction of 50 per cent. Similarly, four predictors—sex, perceived health, bed disability–hospital days, and the respondent's belief in having regular checkups—reduced the prospective difference to 1.2 visits, a reduction of 40 per cent (Table 6).

Concomitant Physical and Psychologic Morbidity

The fact that reported physical morbidity is the best discriminator between users and nonusers of psychiatric facilities suggests that it is important to investigate the link between physical and psychologic morbidity. Our study was not designed to investigate this relationship, but by examining the types of physical conditions differentiating the two groups, we can better understand why physical morbidity is predictive of psychiatric use. There are at least four general hypotheses that can account for the association between physical and mental health: 1) certain people are constitutionally resistant or vulnerable to both physical and mental health problems; 2) psychologic distress may be an etiologic factor in certain physical diseases; 3) the effects of physical disease may lead to or exacerbate the development of psychologic problems; or 4) particular persons may be more sensitive to or aware of both physical and mental states. To investigate the relative plausibility of these different explanations, we examined all the items included in the physical morbidity indices and examined which particular conditions best accounted for the higher distress scores among psychiatric users. With few exceptions, users of the psychiatric clinic were more likely than nonusers in the cross-sectional sample to report psychophysiologic and psychogenic symptoms. The only symptoms psychiatric clinic users did not report significantly more frequently were constipation, high blood pressure and asthma attacks. There was virtually no difference between the two groups in reported frequency of constipation ($p = 0.47$) or high blood pressure ($p = 0.65$), but the difference in reported frequency of asthma was in the expected direction and approached statistical significance ($p = 0.08$). Psychiatric clinic users were also more likely to report numerous ailments, poorer perceived health, and more bed and hospital disability. When chronic health conditions were examined, however, there were significant differences for only eight of the 27 conditions. Those reported significantly more frequently by the psychiatric patients were asthma, repeated attacks of sinus problems, stomach ulcer and other chronic stomach problems, allergies, prostate trouble or repeated urinary tract infections, tremor or shaking, and repeated troubles with back or spine.

These results are at best suggestive, but it is notable that with the exception of prostate trouble and repeated urinary tract infections, all the chronic conditions reported more frequently by patients in the psychiatric sample are conditions frequently viewed as having a psychogenic component. This is not true of most of the conditions for which the groups did not differ.

These findings are consistent with two of the hypotheses mentioned previously: that psychologic factors may be etiologic factors in certain diseases, and that certain persons may be more attentive to both their mental and physical condition. For all the chronic conditions more prevalent in the psychiatric sample, with the exception of prostate and urinary tract problems, one could reasonably argue that either 1) stress could lead to or exacerbate the problem, or 2) that it is the type of condition that would be more likely to be noticed and reported if a person was more sensitive and attentive to his/her physical state. On the other hand, none of the severely disabling conditions, such as stroke, arthritis, diabetes, cancer and paralysis, which might be expected to result in despair or depression, was significantly more prevalent among the users of the psychiatric clinic. It is becoming increasingly recognized that the linkage between physical and mental health is a complex and important one. Understanding the utilization patterns of persons with psychiatric problems will require careful analysis of the occurrence of symptoms.

Multivariate Analyses Within Samples

All of the analyses, thus far, have been confined to differences between groups. As part of our analysis, we also examined factors affecting utilization within the cross-sectional samples of mental health help-seekers and non-help seekers, and within the psychiatric sample. Because the mental health help-seeking samples are small, very few predictors achieve statistical significance, and the only factors we could identify within the psychiatric sample and the help-seeking cross-sectional sample that accounted for differences in medical utilization were measures of health perceptions and disability.

In the case of cross-sectional respondents who did not seek mental health services, significant predictors of use of medical care were access to care, illness behavior and physical symptoms. Whether the respondent is in a prepayment plan, can identify an individual personal doctor and reports little difficulty in getting an appointment help explain medical utilization. These three predictors account for 5 per cent of utilization in the retrospective data, and the first two predictors account for 2 per cent of utilization in the prospective data.

On the assumption that access to care enables people to use services who wish them (either because they have symptoms or particular illness behavior patterns), we entered the three access variables mentioned above as fixed predictors and then examined what combination of physical illness measures and illness behavior measures yielded the best predictors of variations in medical utilization. In the case of prior utilization, the access measures—reported bed-disability days and belief in the importance of regular checkups—accounted for 15 per cent of variation in utilization. The best predictor was reported bed–disability. We were less successful in predicting prospective medical utilization. Membership in a prepaid plan, identifying a personal physician, reported bed–disability days and subjective assessment of physical condition accounted for 6 per cent of variation in utilization (Table 7).

As noted in the introduction, a central issue in the study of medical care use has been whether providing psychologic services reduces the utilization of other medical services by distressed patients. One hypothesis is that high utilization rates may reflect, in part, help seeking for psychologic problems, and if psychologic services are provided they will be "offset" by a corresponding reduction in medical care use. Although our data were not collected to address this issue, by comparing the utilization patterns of distressed patients who used the psychiatric clinic with those of patients in the population sample with comparable levels of distress, we can get some indication as to whether an offset has occurred among the users of the psychiatric clinic.

Because one would expect the impact of psychiatric services to be greatest for those patients who are most distressed, we analyzed utilization patterns among those patients with a demoralization score that

TABLE 7. Regression Models of Medical Care Utilization in the Cross-Sectional Sample

	Retrospective (beta)	Prospective (beta)
Model 1—Access		
Not in prepaid plan	−0.13*	−0.10*
Does not name a regular doctor	−0.13*	−0.08*
Difficulty getting an appointment	−0.09*	
R^2	0.05	0.02
Model 2—Access, Illness Behavior and Physical Symptoms		
Fixed variables		
Not in a prepaid plan	−0.11*	−0.09*
Does not name a regular doctor	−0.04	−0.05
Difficulty getting an appointment	−0.07	
Second step		
Bed–hospital days	0.29*	0.16*
Says regular checkups important	0.12*	
Poor subjective physical condition		0.12*
R^2	0.15	0.06

* $p < .05$.

was above the 90th percentile of the general population sample. Thus, the figures reported are based on 46 people from the sample of psychiatric clinic users and 76 people from the population sample who had demoralization scores above this criterion. Among these people, the change in utilization rates over the period of the study supports the offset hypothesis; the utilization of distressed people who did not use psychologic services increased from 3.04 visits per year (in the year prior to the interview) to 3.25 visits per year (in the year subsequent to the interview), while medical care use among distressed persons from the sample of psychiatric clinic users decreased from 5.54 visits per year to 4.57 visits per year. However, neither of these differences nor the contrast of these differences was statistically significant. In addition, there are several limitations of the data that must be kept in mind when interpreting these results. First, we have no way of determining if any of the medical care visits by people in the psychiatric sample were solely for the purpose of obtaining a referral to the psychiatric clinic. Thus the decrease of less than one visit per year among psychiatric clinic users might partially be accounted for by a referral visit in the year preceding the interview. A second problem is one common to almost all studies addressing the offset question: the sample of psychiatric clinic users is a self-selected sample. Presumably, the persons who seek psychiatric care are most likely to do so at a time when their need for help is greatest. Consequently, one would expect that their use of nonpsychiatric services would also be highest preceding such a visit. Even if there were no treatment effects, one would expect some reduction of both distress and help-seeking behavior (both psychiatric and nonpsychiatric). In short, although there is some indication of an offset effect among patients receiving psychiatric care, this effect, even if free of methodologic ambiguities, would be trivial in comparison with the excess medical utilization of psychiatric patients as compared with patients who are not in psychiatric care.

Discussion

In short, we have been able to explain almost half of the differences in medical utilization among groups who are mental health care seekers and nonusers of mental health services. We have confirmed, using independent records and on a prospective basis, that users of mental health services also use substantially more general medical services. Such differences are in part due to physical symptoms and disorder concomitant with psychologic disorder and higher illness behavior propensities. Physical symptoms and disability appear to have more impact on utilization than differences in illness behavior.

A surprising finding in the cross-sectional sample was that psychologic distress as measured by various screening scales had only a very modest relationship to retrospective and prospective medical utilization. While such relationships are significant at the zero-order level, they are not significant when controls are introduced for access, physical symptoms and illness behavior. The weak impact of distress on utilization relative to some other studies may reflect the special characteristics of this rural population. The population studied appears to be a highly stoical, self-reliant group, many of Northern European and Scandinavian backgrounds. In the United States in 1977, there were 3.3 visits per person to a doctor's office, clinic or group practice, and 3.2 visits in the north central states and outside Standard Metropolitan Statistical Areas.[10] In contrast, our cross-sectional sample had 2.4 visits for the prospective period and 2.6 visits for the retrospective period. It is plausible that distress as a determinant of general medical services is more substantial in situations in which medical care is defined broadly and in which discretionary use is more consistent with cultural patterns in the population.

The data presented here are not consistent with a growing literature suggesting that psychiatric intervention among distressed persons results in reduced medical utilization. Persons receiving mental health services had considerably higher medical utilization for both the retrospective year and for the year following contact with psychiatric services. It is possible that medical utilization for these patients will show a reduction in subsequent years. There are many reasons to assist psychologically distressed persons. Whether reduction in medical utilization occurs has not been firmly established, and studies demonstrating this effect have significant methodologic limitations.

In sum, there are at least three general interpretations for the high rates of medical utilization among patients who receive psychiatric services: that the link is due to the interrelationships between psychologic symptoms and physical symptoms; that it reflects high generalized help-seeking propensities; and that it reflects broader access to both medical and psychiatric services. Although we find some evidence in support of each of these interpretations, the explanation that is most powerful is that high medical utilization among patients receiving mental health services is a product of physical symptoms and discomfort concomitant with psychologic disorder.

References

1. President's Commission on Mental Health. Report to the President, Vol. I. Washington, D.C.: United States Government Printing Office, 1979.
2. Jones K, Vischi T. Impact of alcohol, drug abuse and mental health treatment on medical care utilization: a review of the research literature. Med Care 1979;17(Suppl.): entire issue.
3. Mumford E, Schlesinger H, Glass G. Problems of analyzing the cost offset of including a mental health component in primary care. In: Institute of Medicine. Mental Health Services in General Health Care. Washington, D.C.: National Academy of Sciences, 1979; 101.
4. Hankin J, Okay J. Mental disorder and primary medical care: an analytical review of the literature. Rockville, Maryland: National Institute of Mental Health Series D, No. 5, 1979.
5. Eastwood M. The relation between physical and mental illness. Toronto: University of Toronto Press, 1975.
6. Tessler R, Mechanic D. Psychological distress and perceived health status. J Health Soc Behav 1978;19:254.
7. Tessler R, Mechanic D, Dimond M. The effect of psychological distress on physician utilization: a prospective study. J Health Soc Behav 1976;17:353.
8. Dohrenwend B, et al. Nonspecific psychological distress and other dimensions of psychopathology in the general population: measures for use in the general population. Arch Gen Psychiatry 1980;37:1129.
9. Langner TS. A twenty-two item screening score of psychiatric symptoms indicating impairment. J Health Soc Behav 1962;3:269.
10. Office of Health Research, Statistics and Technology. Health, United States. Hyattsville, Maryland: Public Health Service, 1979. (DHEW Publication (PHS) 80-1232. p. 186.)

MEDICAL CARE
April 1982, Vol. XX, No. 4

Psychological Distress and Patient Satisfaction

JAMES R. GREENLEY, PH.D.,*† THERESA B. YOUNG, M.S.,‡
AND RICHARD A. SCHOENHERR, PH.D.†

Psychologically distressed patients and clients of health care and social service organizations are found to report somewhat more dissatisfaction with services than do the nondistressed. Four explanations for this relationship are examined: 1) the psychologically distressed are generally dissatisfied; 2) service providers react negatively to the psychologically distressed; 3) psychologically distressed patients are dissatisfied when service providers do not respond to their psychological needs; and 4) patients who deny their psychological distress tend to be dissatisfied. The results show that the psychologically distressed report more dissatisfaction because of the very high levels of dissatisfaction found among patients who deny having personal problems.

PERSONS EXPERIENCING higher levels of psychological distress are often thought to be generally less satisfied with medical and related health and social sevices. Reports from medical, rehabilitation, psychological and social welfare settings suggest that attempts to help persons with psychological problems are often difficult and have negative outcomes. Higher levels of psychological distress have been associated with negative provider–client interaction,[1–5] dropping out of treatment programs,[6] poor compliance and refractoriness to treatment in general,[3,7,8] postoperative complications[9] and lengthy convalescence.[10] Despite these numerous accounts linking psychological distress to negative provider–client interaction, recent reviews of patient satisfaction studies show no empirical research linking psychological distress specifically with patient dissatisfaction.[11,12]

Increasing our understanding of the relationship between patient satisfaction and psychological distress may help us better understand why psychologically distressed patients experience such a wide variety of other negative outcomes. Patient satisfaction is, of course, important in its own right, especially in this era of increasing consumerism in medicine.[13]

Despite numerous suggestions and informal observations linking psychological distress to patient dissatisfaction, to our knowledge no prior empirical research has focused specifically on a patient's

---

From the Departments of Psychiatry,* Sociology† and Preventive Medicine,‡ University of Wisconsin-Madison, Madison, Wisconsin.

This research was supported in part by grants from the National Institute of Health (HS-01189) and the National Institute of Mental Health (MH-14641). The authors wish to thank John DeLamater for helpful comments on an earlier draft of this article.

Reprint requests: James R. Greenley, Department of Psychiatry, University of Wisconsin, 600 Highland Avenue, Madison, WI 53792.

0025-7079/82/0400/0373/$01.15 © J. B. Lippincott Co.

psychological distress and satisfaction with service. This article presents research demonstrating that psychologically distressed persons are more dissatisfied in a broad range of patient and client care settings.

Conceptual Background

Several general explanations for this relationship have been suggested in the literature, and these explanations serve to guide our exploration. First, the psychologically distressed are thought to be generally dissatisfied with their lives and experiences, and this is reflected in their assessments of health care and related services. For example, Coleman states that because of fear, anxiety, apprehensions and general life orientation, neurotics, in particular, are prone to be pessimistic and generally dissatisfied.[14] We will call this the "general-dissatisfaction hypothesis."

Second, helping professionals may respond in negative ways to psychologically distressed patients,[2,4,7,15] resulting in patient dissatisfaction. For example, Gunther observed professional rehabilitation staff to have distorted and nonemphatic attitudes toward psychologically distressed patients.[3] Paraprofessional staff with "moralistic" attitudes were observed to want to discipline and reform psychologically difficult patients. Patients with psychological problems may lead physicians to respond with anger, avoidance and frustration, and to refer negatively to them as "crocks" and "hypochondriacs."[1,2,4,7] As a result, psychologically distressed patients may be rejected for service or referred more often than nondistressed helpseekers. Even a well-intentioned referral may leave the patient feeling misunderstood, bitter, hostile, stigmatized or rejected.[16] We will call this the "negative-provider-reaction hypothesis."

Third, dissatisfaction may result from the perceived failure of nonpsychologically oriented professionals to respond to the psychological needs of the helpseeker.[17] Medical patients in particular are often seen as using a somatic presenting problem to legitimate a contact with a physician in which they covertly wish to discuss personal or emotional problems.[18] Patients with such "hidden agendas" may be frustrated and dissatisfied when their physician visit does not provide this opportunity.[2] Other patients and clients may overtly view their physical or social problems as having psychological or personal components, and these persons may be dissatisfied when they do not receive emotional guidance and support from professionals providing services for their physical or social problems.[19] Dissatisfaction, then, may result if the patient or client has overt or covert psychological needs that are not met by the helping professional. This may be called the "unmet-psychological-needs hypothesis."

Fourth, dissatisfaction may more likely be experienced by patients or clients who fail to recognize or deny emotional or personal problems as the source of their distress. Certain distressed individuals avoid defining psychological symptoms as evidence of their emotional or personal problems,[20,21] and externalize the causes of their symptoms and blame others as a defense mechanism,[22] protecting themselves in part from threats to self-esteem and intolerable inner conflicts.[23] Such patients, often said to have "underlying" psychological distress, may select inappropriate help sources, have unrealistic expectations about what a professional can do[1,2] and, particularly, be inclined to blame the provider when their expectations are not met.[3] Patients or clients with underlying or unacknowledged personal or emotional problems may for various reasons be more dissatisfied. We will call this the "denied-psychological-distress hypothesis."

There are certainly other possible hypotheses, but we focus on these because they are the ones most commonly mentioned in the literature. We also recognize that the processes suggested by these four explanations may occur simultaneously and may interact in numerous complex ways. For example, patients or clients who deny or fail to recognize their psychological problems may also be particularly prone to invite negative provider reactions.[4,16] We will try to be sensitive to the possibility of such complex processes in our analysis and discussion, even though we cannot adequately test them all.

Analytic Strategy

Levels of patient dissatisfaction in various psychologically distressed and nondistressed subgroups will be compared to assess the explanatory hypotheses reviewed above. First, for the general-dissatisfaction hypothesis, if the psychologically distressed are generally negative, dissatisfied people, they are expected to express more dissatisfaction, regardless of whether they deny their distress or discuss it with the professional. Second, with regard to the negative-provider-reaction hypothesis, if dissatisfaction is a result of negative provider reactions, then dissatisfied patients should be those who most often experience rejection, referral or other negatively perceived outcomes. Third, for the unmet-psychological-needs hypothesis, if the dissatisfaction is a result of frustration of the psychologically distressed at not having their personal problems dealt with by the professional, then dissatisfaction should be concentrated among the psychologically distressed who, themselves, recognize their problems, but for some reason, do not discuss them with the professional. Fourth, concerning the denied-psychological-distress hypothesis, if dissatisfaction is a consequence of denial of personal and emotional problems, then dissatisfaction should be prominent among the psychologically distressed who report having no personal or emotional problems. Through these comparisons, the nature of the relationship between psychological distress and patient dissatisfaction can be specified to a greater degree.

Methods

Design and Sample

Personal interviews by professional interviewers were obtained with 366 patients and clients of ten health, rehabilitation and related social service agencies.[24] A representative sample of patients or clients seeking help for a new, not continuing or chronic problem were randomly sampled at each agency. An average of 75 percent of patients or clients initially approached consented to be interviewed for the study, and, of these, 75 per cent were successfully interviewed in their homes on an average of six weeks following their initial contact (overall response rate of 56 per cent). While reasons for nonresponse were varied, including inability to locate highly mobile students and others, institutional placement and severely deteriorated physical condition, most nonrespondents refused the interview. Respondents were asked about their sociodemographic characteristics, attitudes, help-seeking experiences, psychological symptoms and satisfaction with the service they had received. The number of respondents from each agency ranged from 26 in one small agency to 45 in two large medical facilities, with an average of 37 respondents per organization.

Over 100 formal help-giving organizations having four or more staff, excluding private practice settings, were identified in the metropolitan area of a medium sized midwestern city. Organizations providing direct service were purposely selected to provide a broad representation by size and type of service; 92 per cent of those invited participated in the study. One organization, a psychiatric inpatient service, was dropped from this analysis as inappropri-

ate, leaving ten organizations. Participating organizations included a general hospital, an adult primary care clinic, a specialty medical clinic, a specialized nursing care service, a public rehabilitation agency, a private rehabilitation agency, a religiously affiliated social service agency, a local public organization advising on specialized public benefits, a social work division of a public agency providing material assistance and a public employment agency. In this diverse group of organizations, persons seeking help are variously referred to as patients or clients, and the providers of service are professionals of different degrees. While for convenience sake, we will sometimes refer to respondents as either patients or clients and to the providers as professionals or not, the reader should remain aware of the diversity of the sample.

Measures

Three measures of satisfaction are used. Two of these are additive indices constructed from questions in which respondents indicated their degree of satisfaction on a 7-point scale ranging from "extremely satisfied" to "extremely dissatisfied." A "humaneness" index includes five items concerning courtesy and friendliness of staff, staff warmth and interest in the client, staff consideration for the patient's or client's feelings, respect accorded by staff and concern expressed by staff for the patient's or client's problem. A four-item "competence" index includes two questions on the general quality of service, one on the technical competence and training of staff, and one on the comprehensiveness of services provided. Cronbach's Alpha for these indices were 0.94 and 0.88, respectively. These indices represent two of the most salient dimensions of patient satisfaction, and patient satisfaction scales of this nature have been shown to be adequately valid and reliable.[11,12] In addition, a one-item rating of the general quality of service provided was used ["How satisfied are you with the general quality of the services you received from (organization)?"] Responses were on a 5-point scale ranging from "very satisfied" to "very dissatisfied."

For our purposes, psychological distress refers to the experience of psychological symptoms of which moderate depression and anxiety are the most common. It does not include experience of those severe symptoms indicative of psychotic states, and may or may not reflect a recognized psychiatric disorder.

To measure psychological distress, the Crandall and Dohrenwend ten-item psychological symptoms cluster of the Langner 22-item psychiatric screening index was used.[25] The Langner index is widely used as a measure of middle-range severity of persistent anxiety and depressive states.[26] The ten-item subindex identified by Crandall and Dohrenwend as indicative of "primarily psychological and rarely organic" problems was used in order to minimize confounds of the Langner index with physical symptoms,[27-29] common among our respondents. These items concern an inability to "get going," feeling apart even among friends, feeling that nothing turns out right, wondering if anything is worthwhile, nervousness, being a "worrying type," restlessness, trouble sleeping, being in low spirits and having a poor memory. An additive index was constructed in which respondents scored one point for each symptom they experienced "often" (Cronbach's Alpha for this index is 0.87).

While psychological distress was measured six weeks following the respondent's approach to the agency for help with the new problem, the Langner items used here measure a sufficiently stable type of psychological distress for this index to provide a satisfactory indicator of the patient's level of psychological distress at the time he or she visited the organization.[30] Test-retest Langner index scores correlate at greater than 0.5 over periods ranging from two to four years,[30–32] and the Langner index correlates more strongly with measures of stable anxiety, neuroticism and depression than with measures of transitory anxiety.[33] In addition, a substantial percentage of the respondents reported still receiving help from the organization at the time of the interview. The only nonrandom bias due to the time lag, which we consider serious, consists of an exacerbation of psychological distress due to a worsening of the problems for which help was sought. To deal with this possibility, two control variables were introduced in all analyses, i.e., the reported degree of worsening of the presenting problem. ["In general, would you say the problem has entirely disappeared or has gotten better, somewhat better, no change, somewhat worse or much worse as a result of your contact with (organization)?] and the degree of worry about the presenting problems at the time of the interview [Since you have been in contact with (organization), are you worrying a great deal, somewhat, very little or not at all about your problem?]. Random fluctuations of psychological distress during the time lag would, of course, conservatively bias against finding any association of satisfaction and psychological distress. Thus, the reasonably short and variable average time lag between the respondent's initial contact with the organization and the interview is unlikely to bias the results.

Psychological distress that is chronic or persistent over periods of months or years may nevertheless provide a trigger or heightened propensity to seek services. Thus, the presence of relatively stable distress levels is consistent with the implication in certain of the alternative explanations that psychological distress may be either the overt or covert reason for which some people seek help. People with higher levels of psychological distress have, for instance, increased use of general medical care over periods as long as a year,[34] and thus the distress probably serves to heighten the propensity to seek help or to reduce the person's tolerance for physical and other problems that would not have been taken for professional help under other circumstances. Thus, people may use services because of or for help with psychological distress, even though the distress levels may be relatively stable.

To examine the negative-provider-reaction hypothesis, patients were asked if they had been "rejected," a clearly negative provider reaction, or "referred," a provider reaction often perceived negatively by patients. Patients were asked, "At (organization) did one or more of the following things happen: you were given some service, that is they tried to help you; you were rejected with no service; you were told to wait or put on a waiting list?" They were also asked, "At any time did (organization) refer you somewhere else or give you the name of somewhere else to go to get help with your problem?" Patients were coded as having been rejected or referred if they so indicated. Referrals were reported by 12 per cent of the respondents and rejections by 3 per cent. Rejection and referral measured in this way were used as separate indicators of negative provider reactions.

Respondents are classified in one of three groups to permit assessment of the other explanatory hypotheses. First, an initial question and related probes were used to elicit respondents' descriptions of all problems they discussed at any time with the service provider. Second, a series of eight questions was used to elicit respondents' descriptions of perceived problems *not* discussed with the provider in each of eight areas: material needs, employment, medical and emotional problems, education, interpersonal problems, leisure time and legal problems. All problems mentioned were categorized as "emotional/personal" by the authors if they were potentially interpersonal or intrapersonal in locus, or if discussion of the problem would involve revealing inner feelings and exposure of the self to others. Using these responses, respondents were placed in one of three comparison groups: 1) those who indicated discussing an emotional/personal problem with the provider; 2) those who reported an emotional/personal problem in response to the series of eight questions described above, but who initially reported discussing no such problems with the provider; and 3) those who reported no emotional/personal problems, either in the initial question concerning discussions with the provider or in the latter questions concerning other problems they had but did not discuss with the provider. Respondents who said they discussed an emotional/personal problem with a provider were considered by definition to have admitted having this problem.

We assumed that respondents would, in general, be willing to discuss personal problems with a research interviewer, and would be at least as willing to discuss these problems with a research interviewer as with a professional, given the stigma often associated with seeking professional help. If respondents did not report personal problems that they themselves recognized, whether or not these were discussed with a provider, those respondents would be incorrectly categorized in group 3 rather than in groups 1 or 2. If this were the case, group 3 would be a more mixed group in terms of problem recognition and presentation to providers, reducing the chances of our finding clear differences between group 3 and the other groups.

Analysis Strategy

Multiple regression analyses are used. As organizations varied greatly in the numbers of clients served, differing proportions of clients contacting the agency were randomly sampled at each organization, to avoid clients from any one organization dominating the data. For a similar reason, our resultant stratified sample is analyzed unweighted, although we recognize existing controversy in this area. Thus, while our sample varies somewhat from a simple random sample, we still find it useful to use statistical significance at the 0.05 level as a standard by which to assess the substantive import of the findings.

Since five respondent characteristics were found to be significantly related to both psychological distress and satisfaction, i.e., sex, age, income, education and adult household role, these variables were included as control variables. For these analyses, age was coded in years, income in 13 categories from under \$2,000 to over \$20,000, and education in years of formal schooling. Major adult household role is a dichotomy between being an adult head of household or spouse of a head of household, and having another household role. These statistical controls guard against the possibility that satisfaction and psychological distress are spuriously related due to these variables.

Results

For the entire sample of 366 respondents, psychological distress is significantly correlated, at the zero-order level, with all three satisfaction measures, i.e., satisfaction with humaneness ($r = -0.10$, $p < 0.05$), competence ($r = -0.22$, $p < 0.05$) and general quality of service ($r = -0.22$, $p < 0.05$). After controlling for the five possibly confounding respondent characteristics described above, psychological distress remains significantly associated with satisfaction with competence and general quality of services in the entire sample.

Several explanatory hypotheses are examined using a separate multiple regression of each of the three satisfaction dependent variables on: psychological distress; dummy variables for two of the nominal categories representing the previously defined subgroups, i.e., the respondent reported personal/emotional problems but did not discuss them with the provider, and the respondent neither reported nor discussed personal problems with the provider (the omitted category was that in which the respondent both reported and discussed personal problems); a multiplicative term representing the interaction of psychological distress and being in the group reporting emotional/personal problems, but not discussing them with a provider; a multiplicative term representing the interaction of psychological distress and being in the group neither reporting nor discussing emotional/personal problems; and the seven control variables discussed previously (see Table 1).

The General-Dissatisfaction Hypotheses

The first explanatory hypothesis is that psychologically distressed patients and clients are generally negative, dissatisfied people. This hypothesis leads us to predict that psychological distress itself would be significantly associated with dissatisfaction, apart from whether the patient admit-

TABLE 1. Regression of Satisfaction

	Dependent Variables		
Independent Variables	Satisfaction With Humaneness	Satisfaction With Competence	Satisfaction With General Quality
Psychologic distress	−0.04	−0.20	−0.06
Control variables			
Sex (female)	0.99	0.71	0.06
Age	0.07*	.05†	.01‡
Education	−0.31†	−0.23‡	−0.01
Income	0.13	0.07	0.00
Major adult household role	1.01‡	0.61	0.22
Worsening of presenting problem	−1.56*	−1.31*	−0.45*
Current worry about presenting problem	−0.31	−0.47	−0.10
Subgroups of total sample			
Admitting personal problems but not discussing them with the provider	−2.35‡	−1.37	−0.24
Not admitting or discussing personal problems with the provider	−1.88	−0.97	−0.03
Interactions of subgroups and psychological distress			
Admitting personal problems but not discussing them with the provider *and* high on psychological distress	0.39	0.54	0.01
Not admitting or discussing personal problems with the provider *and* high on psychological distress	−4.03†	−2.49‡	−1.00‡
R^2 =	−0.17	−0.19	0.16

This table presents unstandardized regression coefficients.
* $p < 0.001$.
† $p < 0.01$.
‡ $p < 0.05$.

ted the distress to himself or herself, or discussed it with a professional. As Table 1 shows, psychological distress itself is not significantly related to the satisfaction dependent variables, but only appears important insofar as it interacts with these other factors. These findings are inconsistent with the first hypotheses.

The Negative-Provider-Reaction Hypothesis

The second hypothesis is that providers react negatively to psychologically distressed patients and clients. Rejection is an overt provider reaction, often indicative of more subtle negative provider attitudes that are difficult to measure. Referrals, however well intentioned, are often experienced by clients as negative reactions of providers. This hypothesis leads us to suspect dissatisfaction among patients and clients who report being rejected or referred. Yet, neither reported rejection nor referral was significantly related to psychological distress or to dissatisfaction in the entire sample or any of the subgroups, and, thus, this hypothesis is not supported.

The Unmet-Psychological-Needs Hypothesis

The third explanatory hypothesis is that patients are dissatisfied when they are not able to discuss or deal with emotional or personal problems they bring openly to providers or have as "hidden agendas." These patients or clients recognize that they have emotional or personal problems, and are thought to be frustrated and dissatisfied at not receiving emotional guidance and support from professionals providing other types of services. If this explanation helps account for the fact that the psychologically distressed report less satisfaction, then the psychologically distressed who admit personal problems but did not discuss them with the provider should be significantly less satisfied. This group of respondents is not less satisfied, as shown by the first interaction term appearing in Table 1, and, thus, this hypothesis does not help us understand the association of psychological distress and dissatisfaction.

Apart from psychological distress, respondents not discussing their personal problems with providers are significantly less satisfied with humaneness of service, although not with competence or general quality of service (see Table 1). This provides at least partial support for the notion that patients with a psychological "hidden agenda," or recognized desires to talk about their problems, may tend to be dissatisfied, making it worthy of further exploration. This finding, however, is independent of the respondent's level of psychological distress.

The Denied-Psychological-Distress Hypothesis

The fourth explanatory hypothesis is that certain psychologically distressed people deny or do not recognize emotional or personal problems, but rather tend to externalize the source of any distress and blame others, including providers of care. Such patients, often said to have "underlying" distress, are thought to have more unrealistic expectations of providers and to be dissatisfied when these expectations are not met. Highly distressed respondents who report no emotional/personal problems are assumed to be denying problems rather than being without problems and nevertheless distressed, a critical assumption addressed more carefully below. This hypothesis predicts that dissatisfaction would be concentrated among the psychologically distressed who do not recognize emotional or personal problems. This group of respondents is significantly less satisfied, as shown by the second interaction term appearing in Table 1. Our findings strongly support the hypothesis that persons who deny or do not recognize their psychological distress tend to be dissatisfied with health care and related services.

Three of the sociodemographic variables were significantly associated with satisfaction on at least one of the three satisfaction measures (see Table 1); higher satisfaction is found among older, less educated patients, as well as among those with major adult household roles. Patients were also significantly less satisfied if they perceived their presenting problems to have worsened since they sought help. Worries about their presenting problems at the time of the interview are not significantly associated with satisfaction.

When the percent dissatisfied are examined (see Table 2), the patterns observed in earlier analyses are replicated. In the total sample, an average of 9.6 per cent indicated dissatisfaction on one of the three satisfaction measures. For those psychologically distressed respondents who do not acknowledge emotional/personal problems, three times as many, or 30.7 per cent, report comparable dissatisfaction. In contrast to most studies of patient or client satisfaction, which show large majorities reporting very high satisfaction and very small percentages indicating dissatisfaction, roughly one third of the psychologically distressed who do not admit emotional/personal problems said that they were dissatisfied.

Finally, the association between psychological distress and dissatisfaction among patients not reporting personal problems was investigated within types of provider organizations to see if the relationship held across the variety of organizations in the sample. The pattern of results was consistent across organizations of different functions, e.g., medical, rehabilitation, social

TABLE 2. Dissatisfaction by Psychological Distress for the Total Sample and Subgroups*

Patient Groups	Psychological Distress (0–2 Symptoms)	Psychological Distress (3–9 Symptoms)	Significance of Difference of Proportions
Total sample	(N = 244)	(N = 122)	
% dissatisfied with humaneness	5	13	0.01
% dissatisfied with competence	7	10	NS
% dissatisfied with general quality	10	20	0.01
Subsamples			
Group A: Acknowledged and discussed personal problems with the provider	(N = 32)	(N = 38)	
% dissatisfied with humaneness	6	3	NS
% dissatisfied with competence	6	3	NS
% dissatisfied with general quality	16	13	NS
Group B: Acknowledged but did not discuss personal problems with provider	(N = 58)	(N = 58)	
% dissatisfied with humaneness	9	10	NS
% dissatisfied with competence	12	9	NS
% dissatisfied with general quality	14	19	NS
Group C: Neither acknowledged nor discussed personal problems with provider	(N = 154)	(N = 26)	
% dissatisfied with humaneness	4	35	0.001
% dissatisfied with competence	6	23	0.002
% dissatisfied with general quality	8	35	0.001

* A respondent is considered dissatisfied if he or she indicated dissatisfaction on the general quality of service question or if the respondent's index scores on the other measures indicated that the average response to the index items were on the dissatisfied end of the continuum.

service or material needs. A wide variety of organization and staff characteristics was investigated in an attempt to locate a factor associated with the presence of psychologically distressed patients or clients not reporting personal problems, but none was found. This suggests that the patterns revealed in this study are general in nature and not confined to organizations with particular functions or other characteristics.

Characteristics of Distressed Patients Not Reporting Personal Problems

We have referred to distressed patients not reporting personal problems as denying them, assuming that most of these patients had emotional/personal problems that they avoided recognizing or reporting. As this assumption is critical to our interpretation, characteristics of the distressed patients not reporting problems were explored to evaluate it. A variety of sociodemographic and attitudinal characteristics of clients was examined to see how psychologically distressed clients not reporting personal problems differed from other psychologically distressed clients.

Among those 122 clients reporting "high" symptom levels (three or more symptoms), not admitting emotional/personal problems was significantly associated with four client opinions: a belief that the presenting problem would get better whether help was sought or not ($r = 0.31, p < 0.001$); a low level of worry about the presenting problem when help was sought ($r = 0.20, p < 0.05$); a low initial estimate of seriousness of the presenting problem ($r = 0.19, p < 0.05$); and a propensity to avoid seeking care of the type sought ($r = 0.22, p < 0.05$). There is no apparent reason that patients should perceive their presenting problem as less serious and troublesome just because they have no other personal problems. In fact, patients having few problems might tend to overemphasize the seriousness of the few problems they do have. Nor is it clear why patients with no personal problems should have negative opinions concerning seeking help. Together, these findings are consistent with a picture of the psychologically distressed who report no personal problems as having a tendency to downplay problems and possibly to avoid becoming dependent on a provider for help. This pattern of results suggests that, as a group, the patients are not free of personal problems, but rather are denying or avoiding them.

Discussion

Small and usually statistically significant relationships between psychological distress and dissatisfaction measures are found using the entire sample. These relationships account for, at most, about 4 per cent of the variance in patient satisfaction and are themselves of questionable practical significance. Subsequent analyses revealed a strikingly dissatisfied group of clients who were psychologically distressed but denied having problems. While this group constituted only 7 per cent of the sample, more than one third of all dissatisfied respondents were in it. A patient in this group of distressed deniers was more than six times as likely to be dissatisfied as someone not in this group, making their experience worthy of considerable attention by researchers and practitioners.

In the exploration of alternative hypotheses that led to this finding, four hypotheses were examined to account for the relationship between psychological distress and satisfaction. First, are psychologically distressed help seekers simply dissatisfied people? The results showed that relationship between psychological distress and satisfaction was not constant across subgroups of distressed respondents differentiated on whether emotional/personal problems were acknowledged or not. Thus, the data do not support this first hypothesis.

Second, is dissatisfaction a result of overt negative provider reactions, which frustrate and alienate the psychologically distressed? Respondent reports of being rejected or referred were used as indicators of provider reactions that might be experienced negatively by patients. No evidence that the psychologically distressed are more likely to experience rejection or referral was found, and neither rejection nor referral could explain the higher dissatisfaction among psychologically distressed help seekers.

Subtle provider attitudes, not translated into the more observable and active responses such as rejection or referral, may be associated with dissatisfaction. As these attitudes were not measured in this study, this possibility cannot be examined. However, if the more overt provider acts, readily visible to clients, are unrelated to satisfaction, then we would not expect the more subtle attitudes, less visible to patients and clients, to be related. In addition, the lack of a relationship is more notable given that even well-motivated referrals of the psychologically distressed are often unfavorably received. This biases the data in favor of finding a negative association between referral and satisfaction, but even so, we did not find such an association.

The third explanation investigated is that dissatisfaction may result when patients feel the need to discuss personal or emotional problems with a professional, but do not have a chance to do so. No support was found for this explanation.

Apart from psychological distress, regression analyses revealed that patients admitting personal problems but not discussing them with a provider were significantly less satisfied with humaneness, but not with competence or general quality of service. This finding is worthy of further research attention. However, the results suggest that the effect of unmet needs to discuss problems on dissatisfaction is largely independent of level of psychological distress.

The major finding of this investigation is that psychologically distressed patients and clients who do not admit emotional or personal problems tend to be markedly more dissatisfied with services. Dissatisfaction tended to be concentrated in this group. We believe that these psychologically distressed persons who did not admit emotional or personal problems were probably denying them. The questionnaires contain multiple questions focusing on various life problems. To be classified as not admitting any personal problems, a respondent would have had to disclaim persistently such problems. Additionally, psychologically distressed persons admitting no personal problems appeared to downplay the seriousness of their problems and to avoid seeking professional help, a pattern consistent with denial. Persons not admitting personal problems, of course, might simply have none. However, this seems more likely among those reporting low rather than high levels of psychological distress. Thus, we feel that the psychologically distressed patients in this study who did not admit personal problems were likely denying or avoiding them.

Denial is a defense mechanism that is common, and often normal and adaptive. It occurs regularly among sufferers from severe illness[35,36] and may be part of a normal recovery process from severe loss. However, denial as conceptualized here is a general personality defense, not denial of a particular problem but rather an avoidance of a wide range of difficulties.

Other types of psychologically disturbed people, not necessarily scoring high on the distress measure used in this study, may also tend to be dissatisfied or to produce negative provider reactions. This would include persons who were passive-dependent and sociopathic, and certain hypochondriacs who use their problems as defenses against unfavorable psychological problems.[37] Thus, this study should not be generalized beyond its particular conceptualization and measurement of psychological distress.

Psychologically distressed persons not admitting to having personal problems were found to have substantially higher levels of dissatisfaction across a range of types of service agencies. The consistency of this pattern across types of agencies and types of client problems suggests that the source of dissatisfaction is associated with individual patient disposition rather than the nature of the providers or the peculiarities of the presenting problems.

The finding of high dissatisfaction among distressed deniers is subject to a variety of explanations. Their dissatisfaction may stem from the processes by which they are selected into care. Their denial or lack of awareness of their problems may result in their more often choosing an inappropriate source of help. The fact that they seek service at all may suggest that they are reluctant help seekers from the beginning, possibly being pushed to seek help by family or friends.

Provider reactions to these distressed deniers may provoke their considerable dissatisfaction. Their general characterization by service providers as more frustrating and less interesting may result in a variety of negative provider reactions, more subtle than the rejection and referral investigated in this study. Within service organizations they may be diverted toward less well-trained and experienced providers; other studies have shown that this happens to less valued clients. These providers may deliver less satisfactory services. Distressed deniers may receive less or different service because they claim fewer and less serious problems. Professionals may believe them, and make less effort and take less interest in their cases, with resultant dissatisfaction.

Finally, distressed deniers may have a general tendency to blame and find fault with others. This could lead them to be less cooperative with treatment suggestions, producing poorer and less satisfying outcomes. Also, these people may tend to blame and have difficulties with family, friends and employers, even if they do not recognize these difficulties.

These explanations—selection into treatment, provider reactions and client characteristics—each have different implications as to what might be the best strategies for care and ways of increasing satisfaction within this group of distressed deniers. For both practical and theoretical reasons, future research should examine these various explanations.

Our sampling procedures give us confidence that the respondents are broadly representative of clients seeking help at these agencies. Whether generalization of the findings to other types of services and communities is justified is another question. While we purposely studied a broad range of agency types, many kinds of services were necessarily omitted, such as clinics treating largely "psychosomatic" disorders, and the processes suggested in this research might not apply in certain settings. In addition, the study was conducted in a city with a large number of specialized and tertiary care services. Clients drawn to and served by such a health and social service system may have different psychological profiles and standards for evaluating their care than do clients in a less resource-rich community. These questions can only be adequately addressed in similar studies of other clients and areas.

While not the major focus of this study, our findings confirm earlier accounts that satisfaction tends to be positively associated with patients' age and negatively associated with years of formal education. Previous research has generally failed to identify a patient group by attitudes or social characteristics that is as dissatisfied as are the patients denying their problems in this study. The relatively high level of dissatisfaction among these patients and clients should be cause for special attention by both researchers and clinicians.

References

1. Abram HS. Psychological responses to illness and hospitalization. Psychosomatics 1969;10:218.

2. Balint M. The doctor, the patient and the illness. New York: International University Press, 1957.

3. Gunther MS. Psychiatric consultation in a rehabilitation hospital: a regression hypothesis. Comprehensive Psychiatry 1971;12:572.

4. Lipsitt DR. Medical and psychological characteristics of "crocks." Int J Psychiatry Med 1970;1:15.

5. Nemiah JC. The psychiatrist and rehabilitation. Arch Phys Med Rehabil 1957;38:143.

6. Calden G, William CL, Thomas HL, Thurston JR. Action research in the psychology of the tuberculosis patient: the problem of irregular discharge. In: Sparer PJ, ed. Personality, stress and tuberculosis. New York: International University Press, 1956.

7. Abram HS. The psychology of chronic illness. J Chron Dis 1972;25:659.

8. Carnes GD. Identifying clients predisposed to failure. Journal of Counseling Psychology 1973;20:79.

9. Janis L. Psychological stress: Psychoanalytic and behavioral studies of surgical patients. New York: Wiley, 1958.

10. Imboden JB, Canter A, Leighton EC. Convalescence from influenza: a study of the psychological and clinical determinants. Arch Intern Med 1961;108:393.

11. Ware JE, Allyson DA, Stewart AL. The measurement and meaning of patient satisfaction. Health and Medical Care Services Review 1978;1:1.

12. Ware JE, Snyder MK, Wright WR. Development and validation of scales to measure patient satisfaction with health care services: Volume I of a final report; Part A, review of literature, overview of methods, and results regarding construction of scale. Carbondale, IL: Southern Illinois University, School of Medicine, 1976.

13. Rivkin MO, Bush PJ. The satisfaction continuum in health care: consumer and provider references. In: Mushkin SJ, ed. Consumer incentives for health care. New York: Prodist, 1974.

14. Coleman JC. Abnormal psychology and modern life. Chicago: Scott, Foresman and Company, 1950.

15. Mastrovito RC. Emotional considerations in cancer and stroke. I. In cancer. NY State J Med 1972;72:1874.

16. Lowy FH. Management of the persistent somatizer. Int J Psychiatry Med 1975;6:227.

17. Mechanic D. Medical sociology. 2nd ed. New York: Free Press, 1978.

18. Shuval JT. Social functions of medical practice. San Francisco: Jossey-Bass, 1970.

19. Weiss RS. Helping relationships: relationships of clients with physicians, social workers, priests, and others. Social Problems 1972;20:319.

20. Mechanic D. Social psychologic factors affecting the presentation of bodily complaints. N Engl J Med 1972;286:1132.

21. Yarrow MR, Schwartz C, Murphy HS. The psychological meaning of mental illness in the family. Journal of Social Issues 1955;11:12.

22. Hoaken P. Paranoid-depressive relationships. Canadian Psychiatric Association Journal 1973; 18:427.

23. Ziegler F, Imboden J, Meyer E. Contemporary conversion reactions: a clinical study. Am J Psychiatry 1970;116:901.

24. Greenley JR, Schoenherr RA. Encounters with Help-Delivery Organizations: Patient and Client experiences. Madison, Wisconsin, Dane County Health and Social Service Agency Project, University of Wisconsin, 1975.

25. Crandell D, Dohrenwend BP. Some relations among psychiatric symptoms, organic illness and social class. Am J Psychiatry 1967;123:1527.

26. Langner TS. A 22-item screening score of psychiatric symptoms indicating impairment. Journal of Health and Human Behavior 1962;3:269.

27. Roberts R, Forthofer R, Fabrega H. Further evidence on dimensionality of the index of psychological stress. Soc Sci Med 1976;10:483.

28. Seiler LH. The 22-item scale used in field studies of mental illness: a question of method, a question of substance, and a question of theory. J Health Soc Behav 1973;14:252.

29. Meile RL. The 22-item index of psychophysiological disorder: psychological or organic symptoms? Soc Sci Med 1972;6:125.

30. Wheaton B. Sociogenesis of psychological disorder: reexamining the causal issues with longitudinal data. American Sociological Review 1978;43:383.

31. Kasl SV, Gore S, Cobb S. The experience of losing a job: reported changes in health, symptoms and illness behavior. Psychosom Med 1975;37:106.

32. Haberman PW. An analysis of retest scores for an index of psychophysiological disturbance. J Health Soc Behav 1965;6:257.

33. Shader RI, Ebert MH, Harmatz JS. Langner's physciatric impairment scale: a short screening device. Am J Psychiatry 1971;128:596.

34. Mechanic D, Cleary P, Greenley JR. Distress syndromes, illness behavior, access to care and medical utilization in a defined population. Forthcoming in Medical Care.

35. Calsyn DA, Louks J, Freeman C. The use of the MMPI with chronic low back pain patients with a mixed diagnosis. J Clin Psychol 1976;32:532.

36. Krystal H, Petty TA. Psychological processes of normal convalescence. Psychosomatics 1961;2(September, October):366.

37. Daniels RS, Margolis P, Carson R. Hospital discharges against medical advice I. Origin and prevention. Arch Gen Psychiatry 1963;8:120.

MEDICAL CARE
April 1982, Vol. XX, No. 4

Life-Styles, Health Status and Medical Care Utilization

CLYDE R. POPE, PH.D.

This study examines the relationship between three life-style variables (drinking, smoking, and physical activity) and three health status measures (physical, mental and general health status) and the use of medical care services. Up to seven years of utilization data were available for linkage with survey data on life-style variables of a random sample of 2,502 adults in an HMO. Relationships between life-style variables and health status measures were examined. Then each life-style variable was examined in relation to medical care utilization for eight disease classes in an ICDA-based disease classification system. A separate analysis is performed for males and females in four age categories. While relationships between the life-style variables and health status measures are not statistically significant for all sex/age groups, they are quite consistently in the expected direction, and life-style variables show some potentially interesting relationships to utilization for diseases with strong psychological or emotional components.

IT IS INCREASINGLY ACCEPTED that health is highly related to life-styles, although the mechanisms responsible for this interaction are not clearly established. Three facets of life-style frequently encountered in the literature on health status and disease etiology are cigarette smoking, drinking of alcoholic beverages and physical exercise

The clearest relationship between a specific morbidity and one highly visible personal behavior is that between cigarette smoking and lung cancer. The association between smoking and heart disease is also quite well established, but there is less certainty that this is an etiologic relationship.

The role of exercise in heart disease has not been clearly established, though data have been reported that are consistent with the hypothesis that physical activity reduces coronary heart disease mortality. Hennekens et al. compared men who died from coronary heart disease with a matched sample of living neighborhood controls and concluded that increased physical activity was associated with a decreased risk of coronary death.[1] Paffenbarger and Hale studied a large group of longshoremen for over 20 years (or to the age of 75 years, or death) and found that

---

Health Services Research Center, Kaiser-Permanente Medical Care Program, Portland, Oregon.

This is a revision of a paper originally presented at the 1978 Annual Meeting of the Pacific Sociological Association, Spokane. The data derive from a study supported in part by a U.S. DHEW grant (No. CH00235) and a Research Scientist Development Award from the National Center for Health Services Research and Development.

0025-7079/82/0400/0402/$01.10 © J. B. Lippincott Co.

men in the high-activity work category, when compared with men in other work categories, experienced a lower coronary death rate, especially in the case of sudden-death syndrome.[2] Morris et al. report data that suggest a lower incidence of coronary heart disease and mortality for workers in more physically active jobs compared with similar workers in less active jobs.[3] A recent study of Harvard male alumni found the risk of first heart attacks to be inversely related to the amount of energy expended through physical activity.[4] Further, exercise was found to be independent of other influences on the risk of heart attack.

There are few data in the literature regarding the effects of physical activity on general health status or on morbidities other than heart disease. There is a general consensus, though, that the health consequences of physical activity are positive.

The effect of alcohol on health has been of longstanding concern. The relationship between a sustained high level of alcohol intake and liver cirrhosis is well established, as is the neurologic outcome of a long history of drinking. Alcohol consumption is also associated with accidents and trauma, though the extent of the relationship remains unknown.

A recent report on alcohol and health reviews the data relating alcohol use to cancer and heart disease.[5] According to that report, alcohol consumption is implicated in the development of certain cancers (especially when combined with the use of tobacco) and in cardiomyopathy. However, the report concludes that longitudinal studies suggest a more complicated relationship between alcohol use and coronary heart disease, including a possible beneficial effect.

In addition to the knowledge that particular life-style factors are significantly implicated in specific morbidities, the notion that health status in general is strongly affected by life-style has wide currency. Little empirical work has been done to test that assumption.

The most recent and probably the best evidence for the effect of life-style on health comes from the Human Population Laboratory Alameda County studies. Wiley and Camacho report data on a sample of Alameda County adults surveyed first in 1965 and again in 1974 that show health practices reported in 1965 to be predictive of health status in 1974.[6] Belloc and Breslow had earlier reported findings of a significant relationship between health practices as reported in 1965 and physical health status as of that time.[7] Follow-up studies with mortality data also showed significant relationships between health practices and mortality rates.[8]

The present study seeks to add knowledge in this area by examining the relationship between three life-style variables and three different health status measures (i.e., physical, mental and general health status) and the use of medical care services for broad categories of morbidities.

Research Setting

The setting for this study is the Oregon Region, Kaiser-Permanente Medical Care Program, a prepaid group practice HMO serving the Portland Standard Metropolitan Statistical Area (SMSA). Five basic characteristics shape the organization of the system: voluntary enrollment; prepayment for comprehensive benefits on a service basis; preventive medical care; integrated, hospital-based health care facilities; and provision of physician services through group medical practice.

This medical care program was established in 1943 and now enrolls more than 250,000 members, or about 20 per cent of the population of the Portland metropolitan area. The demographic and socioeconomic characteristics of the health plan membership correspond closely to those of the Portland SMSA.[9]

Study Population and Data Sources

The population for the present analysis comprises 2,502 adults, a subsample of the population whose medical care utilization first began to be recorded for research purposes in January 1967. This subsample, defined as subscriber units (i.e., families) with at least two years of membership in the Kaiser-Permanente (K-P) program, was extensively interviewed in 1970–1971 to obtain detailed demographic, social and economic data, and data of a situational and attitudinal nature for subsequent linkage to medical record data. Details on the design of the survey and the characteristics of the survey population are reported elsewhere.[10]

The total medical record of this population has been computerized, and all outpatient contacts (including telephone calls and letters) have been recorded from January 1, 1967 through 1973. Data for each contact include time, place, type of service, type of provider, presenting and associated morbidity, symptoms, episodes and procedures indicating the content of the visit. Inpatient data are also recorded for this population.

In the 1970–1971 survey, information also was obtained about the use of physicians outside the Kaiser-Permanente system during the 12 months preceding the interview. The findings, reported in an earlier paper, showed little outside physician use.[11] Hence, it seems likely that the medical care utilization data for this population are inclusive for the periods of time considered.

Measures

The three life-style variables examined in this study are drinking practices, smoking behavior and physical activity, as reported by respondents in the personal interview survey. These are three of the six practices examined in the Alameda County studies. While there are similarities in the way they have been measured, the two studies are not identical because the questions from the surveys upon which the measures are based differ.

Drinking patterns are defined by the quantity of alcohol consumed per occasion combined with the frequency of occasions of drinking.* This quantity–frequency typology is shown in Table 1.

From the typology an index of drinking was created to provide an ordinal measure, grouping together those drinking patterns that are assumed to be most similar in their impact on health. The scores range from 1 to 4 and combine drinking types as follows: 1) abstainers and both regular and occasional light drinkers; 2) occasional heavy drinkers and regular drinkers, occasionally heavy; 3) regular drinkers, frequently heavy; and 4) regular drinkers, heavy. This is a conservative measure of drinking behavior, although it is limited because it includes nothing on the history of drinking. The assumption, therefore, is that current drinking practices reflect past practices. This may be true for most, but it is not likely to be true for all; past heavy drinkers may now be abstainers or light drinkers, for example.

Smoking behavior is measured in terms of smoking history and, among current smokers, by the number of cigarettes smoked. The index scores range from 1 to 9 as given below. Light smokers are defined as smoking half a pack or less per day, medium smokers as one half to one pack per day, and heavy smokers as more than one pack a day. The full index is as follows: 1) never smoked; 2) formerly smoked but stopped more than ten years ago; 3) formerly smoked but stopped within the past ten years; 4) current light smoker, has

---

* This typology derives from a quantity–frequency coding scheme suggested in another connection by Dr. Robin Room, Social Research Group, School of Public Health, University of California, Berkeley. While Dr. Room's contribution is gratefully acknowledged, he is not responsible for the resultant typology.

TABLE 1. Quantity–Frequency Typology of Drinking Practices

	Type of Drinker					
	Regular Drinker				Occasional Drinker	
Quantity–Frequency	Heavy	Frequently Heavy	Occasionally Heavy	Light	Heavy	Light
Occasions on which 6 or more drinks consumed per sitting	At least once weekly	1–3 times monthly	Less than once a month	Never	1 or more times a year	Never
Frequency of drinking	Weekly	Weekly	Weekly	Weekly	Less than weekly	Less than weekly

smoked ten years or less; 5) current light smoker, has smoked more than ten years; 6) current medium smoker, has smoked ten years or less; 7) current medium smoker, has smoked more than ten years; 8) current heavy smoker, has smoked ten years or less; 9) current heavy smoker, has smoked more than ten years.

A physical activity index was created from responses to an open-ended question about the kinds of leisure-time activities engaged in. Six types of activities entailing significant physical exertion were identified and arbitrarily weighted and summed to produce the index. The weighting was arbitrary because the questioning did not ascertain the amount of time spent at the activity per occasion, the number of occasions per interval, and the length of time the activity had been included in the person's repertoire of behavior.

Types of activities that tend to be seasonal and/or do not inherently require a lot of physical exertion were assigned weights of 1, while types that span seasons or are not inherently seasonal and/or do require considerable physical exertion when pursued were given weights of 4. The four types of activities assigned weights of 1 are camping, fishing, gardening and hunting. The two types of activities given weights of 4 are "participating in sports" and "other physical activities," including walking, jogging, running, swimming, and the like. This ordinal measure, then, has a potential range of scores from 0 to 12.

Medical care utilization for this analysis is the 1967–1973 mean annual rate of outpatient contacts (i.e., office visits and telephone/letter contacts) with the Kaiser-Permanente system for disease classes in the Kaiser clinical-behavioral classification system (except pregnancy and its complications and nondisease, i.e., preventive care, refractive errors, and miscellaneous.).

The classification system, based on ICDA codes, was devised for investigating issues relating to medical care utilization. It purports to reflect a clinical reality from the perspective of physicians while grouping together those diseases that, on an a priori conceptual basis, are likely to produce similar behavioral responses in persons with similar sociodemographic and psychosocial characteristics.

The classification system groups the conditions listed in the ICDA into ten classes: 1) pregnancy and its complications (which are excluded from the present analysis); 2) trauma and adverse effects of external causes; 3) nondisease (e.g., preventive care), refractive error and miscellaneous (also excluded from this analysis); 4) diseases generally requiring hospitalization (except psychiatric hospitalization); 5) diseases with a high emotional component (including psychiatric conditions that may result in hospitalization); 6) treatable chronic disease; 7) nontreatable chronic disease; 8) acute microorganism disease; 9) acute nonmicroorganism disease; and 10) symptoms of undiagnosed disease (i.e., physical symptoms but no diagnosable disease state). The classification system has been described in detail by Hurtado and Greenlick.[12]

The number of contacts for each of the disease classes (except pregnancy and nondisease) has been summarized for 1967 through 1973 for each person in the study population. These summaries have then been converted into mean annual rates to take into account differing lengths of membership in the Kaiser-Permanente program. (Fifty-five per cent of the subjects were enrolled for the full seven-year period, while the remainder were enrolled between two and seven years.)

Three nonutilization-based measures of health status are included in this analysis. One is intended as a measure of physical health status, another as a measure of mental health status, and the third as a general or overall health status measure. All three measures derive from personal interview data rather than medical record data, and none is based on reports of utilization behavior.

The physical health measure was developed from responses to questions about recent symptom experiences. Six specific symptoms—dizziness, diarrhea or constipation, skin rashes, persistent cough, lumps or discolored skin, and sore throat or runny nose—were selected because of their physical manifestations and because they also characterize medically significant diseases not generally regarded as mental or emotional.

A physical health index was developed by multiplying the coded frequency with which respondents reported being bothered by each symptom (never = 0, sometimes = 1, often = 2) by a weighting system intended to reflect the seriousness of the symptom. The weights were developed empirically according to the proportion of all respondents who, in a separate question, reported for each symptom whether they thought that a physician ought to be consulted right away. A weight of 1 was assigned for sore throat and for diarrhea or constipation (since 9 per cent and 13 per cent of the respondents, respectively, said a doctor should be consulted right away). Dizziness and skin rashes were each assigned a weight of 4 (38 per cent and 39 per cent, respectively, said a doctor should be consulted right away). Persistent cough and lumps or discolored patches on the skin were both given a weight of 8 (for each, 83 per cent said a doctor should be consulted right away). The weighted scores were then summed across the six symptoms to produce a possible range of scores from 0 to 52.

Mental health status is measured by the 22-item mental health scale developed in the Midtown Manhattan study[13] and included in the Kaiser-Permanente survey. This measure is purported to be a general measure of mental health, although at least one author suggests that it is primarily a measure of psychophysiological symptomatology and general malaise.[14] The index, with a possible range of scores from 0 to 22, was scored in the conventional manner.

The measure of general health status is simply the self-assessments of subjects, an approach commonly used in health surveys. Each person was asked to characterize his/her health "in general" as excellent, good, fair or poor.

Table 2 displays frequency distributions and mean scores on the life-style and health status measures by the sex/age categories created for the analysis. These descriptive data on the number of subjects in the various sex/age categories and their distributions or mean scores on the life-style and health status measures are provided as background for the analysis.

Also included for background are the data presented in Table 3, showing for each disease class the proportion of males and females in each category who had at least one medical contact for a disease or condition included in that class. Except for "Chronic Nontreatable Diseases" and, for males and females under the age of 50 years, "Diseases Generally Requiring Hospitalization," the proportions with contacts for disease/conditions within each disease class are never less than 25 per cent. In fact, for most of the sex/age groups, the proportion with at least one contact during 1967–1973 is in excess of 50 per cent.

Analysis

The analysis is performed by first examining the relationships between the three life-style variables and the three health status measures. Then each of the life-style variables is examined in relation to the use of medical care services for each of eight disease classes in the Kaiser clinical-behavioral disease classification system. The relationships are expressed as Pearson correlation coefficients.

A separate analysis is performed for males and females in each of four age categories: under the age of 35 years, 35–49, 50–64, and 65 and over. This approach is employed as a way of controlling for the relationships of sex and age to the variables of analytic interest rather than simply controlling for them statistically (e.g., by producing partial-correlation coefficients). Thus, the relationships between the life-style variables and the health status measures and the use of medical care services for the various classes of morbidities can be observed for young adult males/females, early middle-age males/females, late middle-age males/females and older males/females.

The analysis is necessarily a limited test of the relationship between the selected life-style variables and health status. The primary reason is that the three measures of health status and two of the three meas-

TABLE 3. Per Cent of Males and Females in Each Age Category with One or More Contacts During 1967–1973 for Morbidities in the Disease Classification System

Disease Class	Males				Females			
	Under 35 (N = 282)	35–49 (N = 362)	50–64 (N = 306)	Over 65 (N = 205)	Under 35 (N = 369)	35–49 (N = 368)	50–64 (N = 364)	Over 65 (N = 246)
Requiring hospitalization	12.4	16.6	31.0	53.2	17.9	19.3	25.3	37.0
With high emotional component	31.9	32.0	30.1	34.1	44.2	45.6	48.9	43.1
Chronic nontreatable	5.7	7.5	6.9	13.7	7.0	10.9	20.0	10.6
Chronic treatable	56.7	74.3	79.4	92.7	74.0	86.1	85.7	90.6
Acute microorganisms	77.3	72.1	67.6	76.6	88.3	80.4	74.4	73.2
Acute nonmicroorganisms	28.4	35.1	42.8	55.1	33.6	41.0	50.0	63.4
Symptoms of undiagnosed disease	58.5	66.8	65.7	74.1	82.4	76.9	77.2	80.1
Trauma	74.5	69.9	62.4	55.6	64.5	64.9	64.8	64.2

ures of life-style (drinking and physical exercise) are for the same point in time. Health status, therefore, might be a *cause* of life-style rather than a *consequence*. That is, people may abstain from drinking or reduce their intake of alcohol and be physically inactive *because* of their health.

Findings

The three health status measures were significantly correlated with each other for both males and females in all age categories. The correlation of the physical health status measure with the mental health scale was 0.359 for males and 0.329 for females, and with the general health status measure, 0.262 for males and 0.209 for females. The correlation between the mental health scale and general health status was 0.356 for males and 0.374 for females. This pattern is essentially the same for the various age categories. These intercorrelations are not so high as to make the measures redundant, nor so low as to suggest that they are measuring totally unrelated phenomena.

The intercorrelations among the life-style variables showed smoking and drinking behavior to be significantly correlated for both sexes (0.330 for males and 0.357 for females) and in all age groups. There was virtually no relationship between smoking and physical activity (−0.035 for males and 0.022 for females). And while drinking behavior was significantly related to physical activity for both males (0.079) and females (0.061) as a whole, that relationship remained significant within sex/age categories only for males and females between the ages of 35 and 49. The relationships, though low, are positive, indicating that those in this middle-age category who engage more in physical activity also tend to drink more (the "19th hole" phenomenon of golf!).

Life-Style and Health Status

The correlations between the three life-style variables and the three health status

TABLE 4. Correlations between the Life-Style Variables and Health Status Measures by Sex and Age

	Males			Females		
	Drinking	Smoking	Physical Activity	Drinking	Smoking	Physical Activity
Physical health						
Under 35	−0.005	−0.039	0.098*	−0.043	−0.066	0.083
35–49	−0.092*	−0.088*	−0.045	−0.045	−0.018	−0.018
50–64	−0.029	−0.129†	0.048	−0.059	−0.105*	0.032
Over 65	−0.097	−0.158†	−0.015	0.035	−0.153†	0.062
Mental health						
Under 35	−0.054	−0.117*	0.082	−0.124	−0.198‡	0.007
35–49	−0.041	−0.057	0.038	−0.011	−0.094*	0.082
50–64	−0.022	−0.053	0.118*	0.007	−0.095*	0.164‡
Over 65	0.000	−0.038	0.104	−0.045	−0.057	0.100
General health						
Under 35	0.010	−0.022	−0.014	−0.108	−0.187‡	0.110*
35–49	0.011	−0.163‡	0.144†	−0.050	−0.078	0.177‡
50–64	−0.089	−0.107*	0.088	−0.009	−0.065	0.103*
Over 65	−0.111	−0.140*	0.243‡	−0.003	−0.093	0.077

* $p \le 0.05$.
† $p \le 0.01$.
‡ $p \le 0.001$.

measures for the various sex/age groups are shown in Table 4.

There is a negative relationship between drinking and the three health status measures for all but a few sex/age groups. The relationships, however, tend to be weak and nonsignificant. The only statistically significant relationship is between drinking and physical health status for males between the ages of 35 and 49: the greater the drinking, the more the experiencing of physical symptoms.

Smoking behavior is negatively related to each of the three health status measures for all sex/age groups. However, for males under the age of 35, the only significant relationship is for mental health status. Males in the other age categories show significant relationships for both physical health and general health status, but not for mental health status.

Among females, the pattern of significant relationships is different. All but the oldest category of females show a significant relationship between smoking and mental health status, while females in the two oldest categories show significant relationships for physical health status. Only the youngest category of females shows a significant relationship between smoking and general health status.

Physical activity is positively related to the health status measures for all but a few sex/age groups, although most are not statistically significant. For each of the male age categories, there is one significant relationship between physical activity and health status. For males under the age of 35, it is for physical health status; for males 50–64 it is for mental health status; and for males 35–49 and 65 and above, it is general health status.

Again, the pattern of significant relationships among females tends to be different, except that for females between the ages of 50 and 64, like males at that age, there is a significant relationship between physical activity and mental health status. For all but the oldest group of females, there is a significant relationship between physical activity and general health status. None of the female age groups shows a significant relationship between physical activity and physical health status.

Overall, the correlations between the life-style variables and the health status measures tend to support the hypothesis of a relationship between life-style and health status generally, as opposed to only specific health conditions or diseases. The next question to be explored in the analysis is whether there is a relationship between the life-style variables and medical treatment for broad disease classes.

Life-Style Utilization

The relationships between drinking behavior and the use of medical care services for the different disease classes are shown in Table 5. For most sex/age categories, the relationships approach zero or are negative. There are some interesting exceptions. For males under the age of 35, there are significant positive relationships between drinking and the rate of medical care contacts for diseases with a high emotional component, for symptoms of undiagnosed disease and for trauma. For females, the only significant positive relationship is between drinking and rate of contacts for diseases generally requiring hospitalization. For males between the ages of 35 and 49, there are significant negative relationships between drinking and rate of contacts for chronic nontreatable diseases and for symptoms of undiagnosed disease. For females in that age category (35–49), there are significant negative relationships between drinking and chronic treatable and acute microorganism disease.

In general, except for males under the age of 35, people who drink more tend to have fewer contacts for medical care except for trauma. While this could mean that people who drink more are healthier, it could also mean that heavier and more frequent drinkers avoid seeking medical care

TABLE 5. Correlation between Drinking Behavior and Rate of Contacts for the Different Disease Classes by Sex and Age

	Requiring Hospitalization	With High Emotional Component	Chronic Nontreatable	Chronic Treatable	Acute Microorganism	Acute Nonmicroorganism	Symptoms of Undiagnosed Disease	Trauma
Males								
Under 35	0.017	0.124	0.008	0.007	0.069	0.016	0.103*	0.142†
35–49	−0.066	0.005	−0.065	−0.158†	−0.053	0.084	−0.116*	−0.046
50–64	−0.045	−0.071	−0.033	−0.040	−0.045	−0.041	−0.060	0.055
Over 65	0.000	0.007	0.063	−0.044	−0.070	0.046	−0.066	−0.046
Females								
Under 35	0.097*	0.032	−0.013	−0.032	0.047	−0.041	0.018	0.040
35–49	−0.016	−0.022	−0.038	−0.102*	−0.091*	−0.041	−0.030	0.067
50–64	0.010	−0.044	0.008	−0.046	−0.024	−0.028	−0.033	0.082
Over 65	−0.047	0.039	0.018	−0.081	−0.016	0.036	−0.001	0.003

* $p \leq 0.05$.
† $p \leq 0.01$.
‡ $p \leq 0.001$.

except for injuries, or that the less healthy have stopped or reduced their drinking.

The relationships between smoking and the use of medical care services for the different disease classes are shown in Table 6. The relationships tend to be small and frequently negative. The significant positive relationships are between smoking and acute microorganism disease, symptoms of undiagnosed disease, and trauma for women under the age of 35, and for women between the ages of 35 and 49 between smoking and diseases with a high emotional component. The one significant relationship among the male age categories is between smoking and chronic treatable diseases for men aged 65 and above.

For all sex/age groups, there is a positive (though mostly nonsignificant) relation-

TABLE 6. Correlation between Smoking Behavior and Rate of Contacts for the Different Disease Classes by Sex and Age

	Requiring Hospitalization	With High Emotional Component	Chronic Nontreatable	Chronic Treatable	Acute Microorganism	Acute Nonmicroorganism	Symptoms of Undiagnosed Disease	Trauma
Males								
Under 35	0.051	0.058	−0.089	0.007	0.057	0.043	0.086	0.046
35–49	−0.063	0.026	−0.060	−0.071	−0.046	0.007	−0.063	0.005
50–64	−0.050	0.041	0.008	−0.018	0.004	−0.055	−0.043	0.021
Over 65	0.028	0.084	−0.077	0.125*	−0.059	−0.026	0.100	−0.065
Females								
Under 35	0.075	0.040	0.026	−0.036	0.098*	0.050	0.101*	0.149†
35–49	−0.052	0.087*	0.040	−0.089*	0.008	0.039	0.021	0.044
50–64	−0.003	0.011	0.064	−0.010	−0.077	−0.047	0.006	0.016
Over 65	−0.022	0.080	−0.072	−0.027	−0.065	0.051	−0.027	0.046

* $p \leq 0.05$.
† $p \leq 0.01$.
‡ $p \leq 0.001$.

ship between smoking and the rate of contacts for diseases with a high emotional component. The hint of another pattern of positive relationships is between smoking and rate of contacts for symptoms of undiagnosed disease and trauma for most of the sex/age groups.

The relationships between physical activity and the use of medical care services for the different disease classes are shown in Table 7. There is a general tendency for higher levels of physical activity to be associated with lower rates of contacts for most disease classes, though only a few of the relationships are statistically significant.

Among males aged 65 and above, there are significant negative relationships between level of physical activity and rate of contacts for disease generally requiring hospitalization, symptoms of undiagnosed disease and trauma. The only other significant relationships (also negative) are between physical activity and chronic treatable diseases for males aged 50–64 and females aged 35–49 and, for males aged 35–49, acute microorganism diseases.

Summary

While the relationships between the life-style variables and the several health status measures are not statistically significant for all sex/age groups, they are quite consistently in the expected direction. That is, smoking is negatively related to each of the three health status measures for all sex/age groups. For males, these relationships are statistically significant mostly on the physical and the general health status measures, whereas for females, the significant relationships tend to be between smoking and the mental health measure.

The findings on drinking behavior are consistent with cultural norms, which may partially explain the finding of a positive relationship between drinking and the general health status measure for males under the age of 50. Heavy and frequently heavy regular drinkers are more than twice as often found among the younger than among the older males in the study. These younger males much more frequently report their health as excellent; they also tend to use fewer medical services.

TABLE 7. Correlation between Physical Activity and Rate of Contacts for the Different Disease Classes by Sex and Age

	Requiring Hospitalization	With High Emotional Component	Chronic Non-treatable	Chronic Treatable	Acute Micro-organism	Acute Non-micro-organism	Symptoms of Undiagnosed Disease	Trauma
Males								
Under 35	−0.018	0.023	0.050	−0.002	−0.022	−0.022	0.026	0.024
35–49	−0.060	0.029	0.069	−0.053	−0.111*	−0.007	−0.055	0.028
50–64	0.011	−0.074	0.021	−0.112*	−0.061	−0.048	−0.088	0.002
Over 65	−0.163†	−0.078	0.060	−0.061	−0.045	−0.050	−0.224‡	−0.161†
Females								
Under 65	0.005	0.009	0.026	0.047	−0.009	0.065	0.015	−0.029
35–49	0.055	−0.065	−0.054	−0.110*	−0.008	−0.008	−0.027	−0.009
50–64	0.024	−0.061	−0.022	−0.027	0.024	−0.013	−0.034	−0.048
Over 65	−0.076	0.024	−0.001	−0.043	−0.039	−0.093	0.039	−0.036

* $p \leq 0.05$.
† $p \leq 0.01$.
‡ $p \leq 0.001$.

The finding of a positive relationship between physical activity and the health status measures for most sex/age groups is probably partially due to the effects of health on activity. The relationships are not strong, however, and only some are statistically significant. Further, neither the physical health measure (based on a limited set of physical symptoms) nor the mental health measure inherently reflects any incapacity for physical activity, yet both measures as well as the general health status measure tend to be positively related to physical activity.

In terms of medical care utilization, the three life-style variables of drinking, smoking and physical activity show some potentially interesting relationships to rate of contacts for diseases or conditions with strong psychological or emotional components. This is the case, however, for only some sex/age groups: for men under the age of 35, it is drinking and medical care contacts for diseases with a high emotional component, symptoms of undiagnosed disease and trauma; for women under the age of 35, it is smoking and contacts for symptoms of undiagnosed disease and trauma; for women aged 35–49 it is smoking and utilization for disease with a high emotional component; for men aged 65 and over, it is (lack of) physical activity and contacts for symptoms of undiagnosed disease and trauma.

To the extent that these findings have meaning, they suggest that younger men with heavier drinking patterns, middle-aged and younger women who smoke more, and older men who engage in few or no physical activities may be candidates for interventions in which the outcome would be reduced utilization for emotion-related disorders. This is a speculation that needs further examination.

This has been a very limited study of the relationship between life-style and general health status and the use of medical care services within broad categories of morbidities. Ultimately, complex longitudinal studies will be necessary to determine the cumulative effects of life-style on general health status and medical care utilization, but there are many less extensive useful studies that can be undertaken. Modifying life-style may be the best or only way of avoiding or containing the personal, social and economic costs of many of today's health problems or diseases.

References

1. Hennekens CH, Rosner B, Jesse MJ, et al. A retrospective study of physical activity and coronary deaths. Int J Epidemiol 1977;6:243.

2. Paffenbarger RS, Hale WE. Work activity and coronary heart mortality. N Engl J Med 1975;292:545.

3. Morris JN, Heady JA, Raffle PAB, et al. Coronary heart-disease and physical activity of work. Lancet 1953;ii:1111.

4. Paffenbarger RS Jr, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. Am J Epidemiol 1978;108:161.

5. U.S. Department of Health, Education, and Welfare, Public Health Service. Second Special Report to the U.S. Congress on Alcohol and Health, from the Secretary of Health, Education and Welfare (preprint edition), 1974.

6. Wiley JA, Camacho TC. Life-style and future health: evidence from the Alameda County study. Prev Med 1980;9:1.

7. Belloc NB, Breslow L. Relationship of physical health status and health practices. Prev Med 1972;1:409.

8. Belloc NB. Relationships of health practices and mortality. Prev Med 1973;2:67.

9. Saward EW, Blank JD, Lamb HB. Some information descriptive of a successfully operating HMO. USDHEW, Public Health Service, Health Services and Mental Health Administration. Washington, D.C.: Government Printing Office, 1973.

10. Pope CR. Data from the 1970-71 Household Interview Survey. Research Report Series No. 1. Portland, Oreg.: Kaiser Foundation Hospitals Health Services Research Center, 1976.

11. Pope CR, Freeborn DK, Greenlick MR. Use of outside physicians by members of a group practice prepayment plan. Paper presented at the Annual Meeting of the American Public Health Association, Atlantic City, November 1972.

12. Hurtado AV, Greenlick MR. A disease classification system for analysis of medical care utilization, with a note on symptom classification. Health Serv Res 1971;6:235.

13. Srole L, Langner TS, Michael ST, et al. Mental health in the metropolis: the Midtown Manhattan Study. Vol 1. New York: McGraw-Hill, 1962.

14. Seiler LH. The 22-item scale used in field studies of mental illness: a question of method, a question of substance, and a question of theory. J Health Soc Behav 1973;14:252.

MEDICAL CARE
April 1982, Vol. XX, No. 4

Stress Potential and Esotericity of Health Problems:

The Significance of the Physician's Affective Behavior

ZEEV BEN-SIRA, PH.D.

The study examines the assumptions and the interpretation of data from recent studies that were intended to identify and explain patients' inclination to assess the efficacy of medical treatment on the basis of physicians' affective behavior toward them. It was hypothesized that this manner of assessment is specific to the doctor–patient relationship due to the uniqueness of the patient's problem, which is characterized by a relatively high stress potential and esotericity requiring professional help. Empirical investigation of four types of problems that varied in the level of their stress potential and the need for professional help ascertains that a) the stress potential of a problem and the need for a professional solution affect *separately* the propensity of assessing the content of the agent's assistance on the basis of the mode of its presentation; b)judgment of the mode indeed determines the assessment of the content of the professional solution; and c) this manner of assessment may apply also to other roles of the medical setting.

THE POSSIBLE EFFECT that patients' satisfaction with medical treatment may have on compliance with medical advice[1] and on recovery[2–6] has been shown in various studies. That, however, raises questions as to patients' capability of assessing the content and efficacy of the treatment and, hence, the factors that determine their satisfaction with it.

Recent studies showed that in a physician–patient interaction—in both ambulatory treatment[7] and hospitalization[8]—the physician's *affective* behavior toward the patient (the physician's *mode* of behavior toward the patient as a human being) will be a major factor in the latter's assessment of the efficacy of the treatment (*content* of behavior). These conclusions were inferred from data that revealed a strong correlation between a layman's evaluation of the physician's treatment (content) and of the practitioner's affective behavior (mode) toward the patient.

From the Hebrew University of Jerusalem, School of Social Work, and the Israel Institute of Applied Social Research, Jerusalem, Israel.

This study is based on data assembled in the Continuing Survey of the Israel Institute of Applied Social Research (at which the author was senior research associate) and the Communications Institute of the Hebrew University. The author expresses gratitude to Professor Aaron Antonovsky of the Ben-Gurion University of the Negev for his assistance and generous advice, and to Mrs. Haya Gratch for her assistance in editing this paper.

Supported by the Israel Institute of Applied Social Research.

Reprint requests: Dr. Zeev Ben-Sira, Baerwald School of Social Work, Hebrew University of Jerusalem, Jerusalem 91004, Israel.

The rationale underlying the interpretation of these data was inferred from the following characteristics of the patient situation: a) patients are emotionally involved as a result of the subjectively perceived seriousness of the disturbance and uncertainty with regard to its consequences; b) they are ignorant of the medical techniques and, hence, incompetent regarding the required solution, in need of professional help, and incapable, in the course of the interaction with the physician, of judging the contribution of the latter's (content) in promoting their recovery. Moreover, frequently, upon completion of the interaction, they will be unable to judge whether there has occurred a change in their condition and the extent to which such a change is attributable to the physician's action.

These characteristics were inferred from a variety of studies that dealt with the patient situation[11,12] and alluded to the significance of the emotional factor in patients' evaluation of the efficacy of their treatment.[6,9,12,14,17]

However, that these characteristics determine patients' inclination to assess the content of the physician's actions (namely the instrumental–technical dimensions of the treatment) on the basis of the mode of their presentation (the affective behavior toward the patients) was merely an assumption. The question of whether this assumption reflects reality is crucial for the validity of that model. The aim of the present study is to investigate whether this assumption is empirically supported.

Specificity of the Model

The assumption underlying that model was that the inclination to assess the instrumental dimension (content) of an agent's response on the basis of the affective dimension (mode of its presentation) is specific to the patient–physician interaction. This inclination, it has been argued, results from the uniqueness of the patient's situation, in which the problem is characterized, more than in any other agent–client situation, by a high degree of a) stress potential and b) esotericity, hence requiring professional solutions.[18,19] Consequently, validation of the model requires that this manner of assessment will apply solely to such situations, and not to circumstances in which clients' problems are not characterized by emotional involvement and not require professional help. To illustrate, we would expect highly worried ill persons to assess the quality of their physician's treatment on the basis of the latter's affective response. Yet, we would not have such expectation with regard to clients' assessment of, say, a plumber's performance, the efficiacy of which may be readily judged, and clients' physical well-being not involved. Thus, in formal, testable terms, we think that empirical support of the following hypothesis is crucial for the validation of the above mentioned model[7,8]: 1. The correlation between clients' satisfaction with the content of their agent's solution, and their satisfaction with the mode of its presentation, will be stronger in a) high stress-potential problems and b) problems requiring a physician's solution.

Satisfaction With the Affective Dimension of the Response as a Determinant of the Satisfaction With the Instrumental Dimension

At this stage, we would like to note that the proposition underlying the model of the professional–client interaction[7,8] is not simply that of a correlation between the two aspects of the assessment. It implies, rather, that in an interaction involving a high level of stress potential and requiring solution by a physician, the satisfaction with the latter's instrumental response (content of solution) will be dependent on the satisfaction with the affective dimension of the response (mode of presentation). In other words, patients' satisfaction with the mode of the agent's response should be, to a great extent, a sufficient condition for their satisfaction with the content of the agent's action. A prediction in the reverse direction would disagree with the rationale of that model. In other words, the rationale of the model implies that when patients are satisfied with the mode of the physician's behavior, there are good chances that they will also be satisfied with the content of the physician's solution. However, the chances of being satisfied with the mode when satisfied with the content are much lower. This asymmetry in the predictions agrees with the evidence that solution of patients' physical problems will not necessarily result in the solution of their emotional problems.[14,17] On the other hand, it is inherent in the logic of the model that laymen's judgment of the quality of their treatment is to a great extent based on the physician's affective behavior toward them (cf. 6, 9, 17). According to the logic of the model, no such asymmetry should be expected when relating to other types of problems, which do not involve a high level of stress potential and which do not require a professional's solution.

Thus, it seems crucial for the validation of the model that the above mentioned asymmetry between the frequency distributions of the two dimensions of satisfaction is confined to the patient–physician interaction when problems are commonly recognized as having a high stress potential. In terms of a testable formulation, the validation of the model also requires empirical support of the following hypothesis: 2. The proportion of patients satisfied with the physician's instrumental (content) behavior among those satisfied with his/her affective behavior (mode) is greater than the proportion of those satisfied with the mode among those satisfied with the content. No such asymmetry of the frequency distribution should be expected with respect to nonstressful problems requiring solution by persons other than physicians.

Method

Sample and Fieldwork

A representative random sample of Israeli Jewish urban adults (N = 492)* was interviewed, using a closed questionnaire presented to them by interviewers in the respondents' homes during January 1977.

Measures

Level of Perceived Stress Potential of Problems. In order to arrive at a scale of the perceived stress potential of problems, the respondents were presented with several types of problems (to be outlined later) and asked whether they had ever faced such problems. If their response was positive,† the stress potential of this problem was measured by four variables that reflected the characteristics of the patient–physician interaction as assumed in the models of previous studies[7,8]:

1. The perceived *seriousness* of the problem (1, very serious . . . 4, not serious at all);
2. The respondent's capability of solving the problem (1, very clear how to solve the problem . . . 4, not clear at all).
3. The salience of the problem, namely, the thought given to it (1, thought about it very much . . . 4, not at all). Since everyone is faced with "demands,"[21] but one pays attention to some more than to others,[22] this variable was intended to measure the attention given to that specific problem. Based on recent approaches of Stress[20] and Breakdown,[21] we considered this variable as essential for measuring the stress potential of the problem, believing that it is not sufficient for a problem to be stress arous-

---

* The sample has been drawn systematically from the Israeli voting register, which is the relatively most reliable updated listing of the Israeli population.

† Only respondents who have faced such problems were investigated, and the results of this study refer to these (See Table 1).

ing by merely considering it as serious and insoluble.

4. The extent to which the respondent was worried by this problem (1, very much . . . 4, not at all) was intended to measure the extent to which an individual feels threatened by this problem, a fact considered to be one of the conditions required for a situation having a stress potential.[22]

Guttman scales, consisting of these four variables for each type of problem, were formed after the replies to each item had been dichotomized (by cutting above and below the median).‡ Thus a scale of the level of stress potential of each type of problem was formed having five degrees: 1, indicating a "very low" stress potential (respondent defining the problem "low" on all four measures), to 5, a "very high" stress potential (problem defined to be "high" on all four variables). With the help of these scales, the types of problems were defined as being of high or low stress potential. The scale served also to differentiate the attitudes of subgroups toward the same problem.

Type of Problem Solver. For the purpose of the study we differentiate between two types: professional (namely, physician) and nonprofessional (e.g., artisan, administrative staff, etc.). This classification served as a basis for the selection of the problems.

Types of Problems. Following the line of our analyses, the types of problems were chosen by two criteria: 1) a hypothetical level of stress potential (to be verified in the course of the analysis on the basis of the responses to the earlier mentioned questions) and 2) the occupation of the required problem solver (physician or nonprofessional). In order to produce a common denominator, only types of problems that might affect respondents or their immediate families' physical well-being were selected. Two types of problems required solution by physicians, one of them by a general practitioner and the other by hospitalization. Considering that hospitalization will usually occur after a GP's intervention, we hypothesized most cases requiring treatment by a GP to be less serious (hence of a lower stress potential) than those requiring hospitalization.

The other two types of problems required solutions by those commonly referred to as nonprofessionals: artisans and administrators. One such problem referred to damages that could be dangerous to the health or life of respondents or their immediate families, and usually required the help of an artisan—such as leakage of gas, uncovered electric wires or damage to the car brakes.§ The second problem concerned health problems of respondents or of their immediate families requiring the help of a medical administrator—such as provision of hospitalization or other medical care. We hypothesized the latter problem to be of higher stress potential, since it was concerned with a threatening event that had already hurt individuals or their families (such as disease), whereas the former is concerned with an event that had occurred but had not yet affected respondents' health although it might be damaging in the future. Thus, four problems were selected and classified according to their empirical level of stress potential, as indicated in Table 1. The percentages presented in Table 1 support empirically the hypothesized level of stress potential. It appears, however, that the type of problems requiring a medical administrator's solution constitute a borderline case. Nevertheless, compared with the other type of problems also requiring

---

‡ The COR's of each of the scales were 0.82, 0.86, 0.88 and 0.78. Though these COR's are lower than the 0.90 as required for a perfect scale, we considered it as sufficient for our purposes, intended in fact to measure the number of variables on which a respondent is "high."

§ Respondents were presented with illustrations of a variety of problems and not with one specific problem.

TABLE 2. Distributions and Mean Scores for Life-Style and Health Status Variables by Sex/Age Groups*

	Males				Females			
Life-Style Variables	Under 35 (N = 282)	35–49 (N = 362)	50–64 (N = 306)	Over 65 (N = 205)	Under 35 (N = 369)	35–49 (N = 368)	50–64 (N = 364)	Over 65 (N = 246)
Drinking index								
1 (Abstainers; light drinkers)	28.0%	42.1%	62.1%	82.0%	60.9%	65.8%	81.5%	92.5%
2	35.1%	26.6%	18.3%	9.3%	32.4%	22.8%	11.3%	5.0%
3	22.7%	13.2%	8.4%	3.6%	3.6%	5.7%	2.0%	1.7%
4 (Heavy regular drinkers)	14.2%	18.1%	11.2%	5.1%	3.1%	5.7%	5.2%	0.8%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Smoking index								
1 (Nonsmoker)	29.6%	22.7%	20.7%	34.8%	49.5%	43.9%	51.5%	77.0%
2	2.1%	10.0%	17.7%	26.4%	1.1%	4.9%	5.3%	6.6%
3	21.8%	27.4%	21.0%	17.9%	16.9%	10.4%	13.0%	5.8%
4	17.9%	2.2%	1.0%	—	18.0%	4.1%	2.0%	1.2%
5	13.2%	21.9%	25.5%	15.4%	9.6%	26.0%	21.3%	7.8%
6	3.9%	—	—	—	1.6%	—	0.3%	—
7	9.3%	13.3%	11.5%	5.5%	1.9%	9.6%	5.5%	1.6%
8	0.4%	—	—	—	0.3%	—	—	—
9 (Heavy, long-term smoker	1.8%	2.5%	2.6%	—	1.1%	1.1%	1.1%	—
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Physical activity index								
Mean score	4.4	4.6	3.6	2.7	3.4	3.1	2.3	1.7
Standard deviation	2.9	3.2	2.7	2.3	2.9	2.9	2.7	2.1
Health Status Variables†								
Physical health index								
Mean score	3.6	3.3	4.6	5.1	5.4	4.3	4.9	4.9
Standard deviation	5.3	5.3	6.7	5.9	6.0	5.2	6.4	5.8
Mental health index								
Mean score	1.8	1.4	1.5	1.7	2.8	2.5	2.8	2.6
Standard deviation	2.1	1.8	2.0	2.1	2.7	2.3	2.6	2.7
General health index								
Mean score	1.6	1.7	2.0	2.2	1.6	1.8	2.1	2.1
Standard deviation	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9

* The numbers (Ns) shown in the table are approximate. For each variable, those few persons with missing data were excluded from the tabulation.

† For each index, the lower the score the better the health. In subsequent tables, however, signs of coefficients are positive when the direction is toward better health and negative when the direction is toward poorer health.

TABLE 1. Classification of Types of Problems by Level of Stress Potential and Type of Problem Solver
(Number in parentheses indicates number of individuals who faced such a problem and turned for a solution to the appropriate role incumbent.*)

Hypothetical and Actual† Level of Stress Potential	Type of Problem Solver			
	Physician		Nonprofessional	
	Per Cent Defining Problem to Have a High‡ Stress Potential	Type of Solution	Per Cent Defining Problem to Have a High‡ Stress Potential	Type of Solution
High†	65 (219)	Treatment by physician in hospital	40 (248)	Solution by medical administrator
Low†	34 (438)	Treatment by GP	22 (209)	Repair by artisan

* Only respondents who faced such a problem and turned for its solution to the appropriate role incumbent were included.

† The "actual" level of stress potential is determined by the percentage defining the problem as high on stress potential and by using the average percentage ($\bar{x} = 40$ per cent) as cutting point: percentages equal or above average were defined as indicating a high level of stress potential.

‡ High = being high on at least three of four items on the scale of stress potential.

low-esotericity solutions, their level of stress potential is much greater.

Satisfaction With Instrumental Behavior (Content). Respondents' assessment of the extent to which the agent's instrumental behavior (content) furthered the achievement of their goals was obtained by asking about the extent of satisfaction with the treatment given to each of the problems by the relevant problem solver (1, very satisfied . . . 6, very dissatisfied).

Satisfaction With Affective Behavior. Respondents' evaluation of the problem solver's affective behavior (mode) was obtained by asking about their satisfaction with the attitude toward them during the interaction. (How satisfied were you with artisan's/GP's, etc., attitude toward you: 1, very satisfied . . . 6, very dissatisfied).

Results

Association Between Satisfaction With Instrumental and With Affective Responses

The data in Table 2 support hypothesis 1. We shall first consider the results related to the total population. The correlation[||] between satisfaction with the instrumental and with the affective response rises with the rise of the level of both the perceived stress potential of the problem and of the professionalization of the problem solver: the monotonic correlation (μ_2) increases gradually from 0.75 with regard

---

[||] A weak monotonicity coefficient (μ_2) was used. This coefficient is especially appropriate in conditions such as ours, in which linearity is not assumed, as there is no assumption of an equal distance between each of the categories in each of the items and the number of effective categories differs from item to item. Since the weak monotonicity coefficient (μ_2) does not assume a precise linear regression, its absolute value will be higher than that of Pearson's coefficient.[23,24] The formula for μ_2 is as follows: given n pairs of observations on numerical variables (x, y), (x_i, y_i) ($i = 1, 2, \ldots, n$), then

$$\mu_2 = \frac{\sum_{h=1}^{n} \sum_{i=1}^{n} (x_h - x_i)(y_h - y_i)}{\sum_{h=1}^{n} \sum_{i=1}^{n} |x_h - x_i|\,|y_h - y_i|}$$

TABLE 2. Monotonic Correlation Coefficients (μ_2) Between Satisfaction With Instrumental and Affective Response to Four Types of Problems by Level of Perceived Stress Potential of the Problems and Type of Problem Solver

Level of Perceived Stress Potential of Problem[*]	Type of Problem Solver	Coefficients of Monotonic Correlation (μ_2) Between Satisfaction With Instrumental and With Affective Response of Agent		
		Total Population	Subgroups Who Perceive Relative Stress Potential as:	
			Low	High
A Low (22%)	Nonprofessional[†]	0.75	0.76	0.74
B Low (34%)	Physician[‡]	0.81	0.82	0.80
C High (40%)	Nonprofessional[§]	0.89	0.84	0.93
D High (65%)	Physician[‖]	0.94	0.83	0.97

[*] Numbers in parentheses indicate percentage of total population who define problem as high on stress potential (see Table 1).

[†] Problem requiring solution by artisan.

[‡] Problem requiring solution by GP.

[§] Problem requiring solution by medical administrator.

[‖] Problem requiring solution by physician in hospital.

to a problem defined as low by both criteria to 0.94 regarding a problem characterized as high on both. Yet, surprisingly, the data apparently indicate greater weight of the stress potential of the problem than of the professionalization of the problem solver in assessing the instrumental behavior (content) on the basis of the affective behavior (mode). This may be inferred from the stronger correlation existing between the two behaviors with respect to the problem requiring a solution by a medical administrator (high stress potential–nonprofessional) than with respect to that requiring a GP (low stress potential–professional). In fact, it appears that both stress potential and the type of problem solver affect, independently, the correlation between the two assesment components.

The hypothesis that both the professionalization of the problem solver and the perceived stress potential of the problem determine the strength of the association gains additional support by comparison of the results among subgroups who consider the problem relatively high on stress potential. Comparison of the data in the two righthand columns (subgroups) for the first two rows (A and B) reveals an increase in the strength of the correlation with the rise of the professionalization of the problem solver, independent of the level of stress potential assigned by the subgroups.

Comparison of rows A and C reveals an increase in the strength of the correlation with the rise of the level of the perceived stress potential of the problem in both subgroups, though the degree of professionalization is low. But when the generally accepted level of stress potential is relatively high (C and D), the increase in association due to the rise in professionalization occurs only for those who assign the problem a relatively high level of stress potential.

Determination of Satisfaction With the Instrumental Behavior of the Problem Solver by the Extent of Satisfaction With Affective Behavior

According to hypothesis 2, the chances for those satisfied with the affective behavior of a physician to be also satisfied with the physician's instrumental behavior are greater than for those satisfied with the instrumental behavior to be also satisfied with the affective. It was also maintained that no such trend would occur with respect to problems that do not involve a high stress potential and do not require a physician's solution.

In order to investigate this issue, we shall employ a simple technique that has recently been demonstrated by Guttman.[25] This technique requires a comparison between both directions of a cross-tabulation between a pair of variables. The conclusions are inferred from the extent of *similarity* of the marginal distribution in either direction. Hence, in order to reach a conclusion as to whether the satisfaction with instrumental behavior of the response is determined by the affective behavior, we shall use cross-tabulations between satisfaction with the instrumental behavior and satisfaction with the affective behavior. The pairwise frequency distributions are presented in Table 3.

Table 3 contains eight subtables, each of the four topics appearing twice in order to show the conditional distribution in each direction; first, the level of satisfaction with the affective behavior is held constant, and then (on the righthand side) the level of satisfaction with the instrumental behavior is held constant. The distinctive feature is that regarding the first problem (low level of stress potential and nonprofessional problem solver), the distributions in either diretion are similar, whereas for each of the other three problems, the distribution differs in each direction. Let us inspect the second problem, which requires a GP's solution (low stress potential and professional problem solver): 84 per cent of the respondents who were "very satisfied" with the affective response were also "very satisfied" with the instrumental response, but almost none (1 per cent) was "not so satisfied," "not satisfied" or "very dissatisfied" with the instrumental behavior. But, on the other hand, only 68 per cent of the "very satisfied" with the problem solver's instrumental behavior were also "very satisfied" with the affective response. Moreover, 91 per cent of the respondents who were "satisfied" with the affective response were also "very satisfied" (19 per cent) or "satisfied" (72 per cent) with the instrumental response, whereas only 55 per cent of those "satisfied" with the instrumental response were also "very satisfied" (9 per cent) or "satisfied" (46 per cent) with the affective response.

An identical trend can also be observed with respect to the fourth problem (hospitalization): 89 per cent, compared with 77 per cent with respect to the "very satisfied"; 96 per cent compared with 84 per cent regarding the "satisfied." Interestingly, however, the third problem (medical administrator) also reveals a similar trend in spite of the nonprofessional solution, yet the high stress potential.

These relationships suggest that when the level of stress potential of a problem is high and/or requires a professional's solution, satisfaction with the problem solver's instrumental behavior is more determined by than coincidental with satisfaction with the affective behavior. Hence, it is suggested to view this type of relationship as indicating that satisfaction with the affective behavior is rather frequently a sufficient condition for satisfaction with the instrumental behavior. If respondents express strong feelings of satisfaction with their agent's affective response, we can predict with a high degree of confidence that they will also be very satisfied (or at least satisfied) with his/her instrumental behavior, but certainly not dissatisfied. However, if they are satisfied with the instrumental behavior, we cannot be sure that they will also be satisified with the affective behavior. These results hold only when the level of perceived stress potential of a problem is high and/or requires a professional's solution—each of

TABLE 3. Mutual Conditional Distributions of Satisfaction with Instrumental and Affective Response to Four Problems by Level of Perceived Stress Potential of Problem and Type of Problem Solver (In Per Cents)

Level of Perceived Stress Potential of the Problem	Type of Problem Solver	Satisfaction With Affective Response	Satisfaction with Instrumental Response							Satisfaction with Instrumental Response					
			Very Satis-fied	Satis-fied	Some-what Satis-fied	Not so Satis-fied	Not Satis-fied	Very Dis-Satis-fied	Total	Very Satis-fied	Satis-fied	Some-what Satis-fied	Not so Satis-fied	Not Satis-fied	Very Dis-satis-fied
Low (problem requiring solution by *artisan*)	Nonprofessional	Very satisfied	65	26	7		2*		100 (43)	62	12	7		4*	
		Satisfied	14	66	14		5		100 (83)	27	61	27		13	
		Somewhat satisfied	5	25	60		10		100 (40)	4	11	55		13	
		Not satisfied	7	33	12		49		100 (43)	7	16	11		70	
										100 (45)	100 (90)	100 (44)		100 (30)	
Low (Problem requiring solution by GP)	Physician	Very satisfied	84	13		2†		1‡	100 (96)	68	9		5†		2‡
		Satisfied	19	72		4		5	100 (87)	14	46		8		9
		Somewhat satisfied	18	35		39		8	100 (66)	10	17		67		11
		Not so satisfied	9	38		17		36	100 (47)	4	13		21		36
		Not satisfied	12	42		—		47	100 (43)	4	14		—		43
										100 (118)	100 (135)		100 (39)		100 (47)
High (Problem requiring solution by medical administrator)	Nonprofessional	Very satisfied	80	17	—	3		‡	100 (30)	63	5	—	3		—
		Satisfied	12	67	6	5		1	100 (91)	31	72	13	13		3
		Somewhat satisfied	—	16	61	16		7	100 (57)	—	10	74	24		12
		Not so satisfied	3	19	6	59		13	100 (32)						
		Not satisfied and	2	15	10	10		63	100 (40)	3	6	4	50		12
		very dissatisfied‡								3	6	9	11		74
										100 (38)	100 (93)	100 (47)	100 (38)		100 (34)
High (Problem requiring solution by physician in hospital)	Physician	Very satisfied	89	10	—		1*		100 (100)	77	13	—			5*
		Satisfied	29	67	1		3		100 (79)	20	71	9			11
		Somewhat satisfied	4	35	43		17		100 (23)	1	11	91			21
		Not satisfied*	11	22	—		67		100 (18)	2	5	—			63
										(115)	(75)	(11)			(19)

* Due to size of N, categories "not so satisfied," "not satisfied," "very dissatisfied" were collapsed.

† Due to size of N, "somewhat satisfied" and "not so satisfied" were collapsed.

‡ Due to size of N, "not satisfied" and "very dissatisfied" were collapsed.

these factors apparently having separate significance. In fact, when both the stress potential of a problem and the professionalization of the problem solver are high (e.g., hospitalization), satisfaction with the agent's instrumental response will be determined, more often than in any other situation, by the agent's affective response: 89 per cent of the very satisfied with the affective behavior are very satisfied with the instrumental behavior, compared with 84 per cent (in the case of the GP) and 80 per cent (in the case of medical administrator). Furthermore, almost all 94 per cent of the "satisfied" with the affective behavior are very satisfied (29 per cent) or satisfied (67 per cent) with the instrumental behavior.

We think that these results support these suggestions by showing that when people face problems that are characterized by a high level of stress potential and that require solution by a professional, their satisfaction with the agent's instrumental behavior will frequently be determined by their satisfaction with the agent's affective behavior. These results seem to indicate that stress potential and the problem solver's professionalization affect *independently* the inclination to assess the agent's instrumental behavior on the basis of his/her affective behavior. That seems to be, to some extent, not entirely compatible with the logic by which both stress potential and professionalization are conceived to be necessary factors for such a manner of evaluation. We shall relate to this question in the discussion below.

Discussion and Summary

In this study, an attempt was made to investigate the extent to which the assumptions underlying a previously published model of professional–client relationship indeed determine the conclusions that were reached in those studies.[7,8] The assumptions that were considered crucial for the validity of that model were as follows: Clients' (patients' satisfaction with the content of the professional's (physician's) activities (medical treatment) will, to a great extent, be determined by the mode of their presentation, when clients (patients) a) are emotionally involved in the problem requiring the professional's (GP's) solution; b) in the course of the interaction lack the skills to judge the extent to which the professional's (physician's) activities contribute to their goal achievement; and c) frequently, upon completion of the interaction, are unable to attribute the goal achievement to the professional's (physician's) activities.

In order to validate the assumptions underlying that model, it was hypothesized that when clients face problems that involve a high stress potential and require a professional's solution, their satisfaction with the professional's instrumental behavior will be determined, in large measure, by the mode (affective behavior) of the response. Results of the study supported the hypothesis. However, data indicated that in contrast to the hypothesized trend, the stress potential of the problem and the need for a professional's solution have an *independent* impact on the extent to which clients' satisfaction with the instrumental component of the response will be determined by the affective behavior. Does that disagree with assumptions underlying the said patient–physician interaction model?[7,8] On the one hand, following strictly the model assumptions, it appears that the underlying logic requires the joint effect of both the stress potential and the need for professional help. On the other hand, however, it can be argued that the model did not specify the weight of each factor. Hence, our results may be considered supportive of that model. The data allude to the weight that each factor may have in enhancing the ability to judge the instrumental response with the help of the affective. Consequently, the greatest chances for judging the instrumental response in such a manner will be when the problem involves both a high stress potential and professional help. Yet, the data regarding the case of the medical administrator are noteworthy. This "problem solver" occupies a somewhat intermediate position between the GP and hospitalization. Does this allude to the greater weight of the stress potential as compared with that of the need for professional help? This would contradict the logic of the model. The assumption was that patients apply the mode as a measure because of their ignorance of the medical techniques. That, however, is not the case with respect to the medical administrator, who applies clerical techniqies that should be considered judgeable by a layman. Two possible explanations are offered. On the one hand, we may safely assume that in a majority of circumstances, particularly in the Israeli setting where medical services are almost entirely prepaid, the medical administrator controls vital medical resources that frequently may even outrank those of the physician. Among these resources should be mentioned giving or withholding information, which is vital for clients who lack the proficiency for checking its correctness (such as the length of the waiting list for some vital medical service, available vacancies in maternity wards, etc.) or making decisions regarding referrals to medical institutions, including priorities, etc. On the other hand, it could be argued that since most services required from a medical administrator involve a high stress potential (as indeed the data in Table 1 indicate), and since frequently no immediate solution is attainable (a long waiting list for a specialist, no vacancy in preferred institution, etc.), the medical administrator's explanations and encouragement may indeed fulfill an important emotionally supportive function. In fact, we maintain that in such circumstances the situation of the administrator resembles to some extent that of a physician, namely, clients' ignorance regarding the required solution and the administrator's solution to a vital stressful problem. Under such circumstances, the emotional support given by the administrator may have a tension-alleviating function and consequently enhance clients' well-being. This conclusion seems to be congruent with the studies that showed the importance of the emotional support given to patients by the hospital staff in furthering their recovery.[4–6,9,17]

Thus, we think that our data support the previous model of the patient–physician interaction[7,8] by proving the significance of emotional support. However, they highlight an additional point by indicating the significance of emotional support provided not only by the physician, but also by those auxiliary personnel who are involved in solving highly pressing health problems.

Additional research will be required to verify these conclusions and their applicability to other cultural contexts and other situations, specifically those in which the help of auxiliary medical personnel is involved. We believe, however, that the significance of the present study is in its clarification of the utility of the model of the physician–patient interaction[7,8] by showing a) its specificity to the situation of the medical setting, proving that clients' propensity to assess the quality of an agent's activity on the basis of the mode of presentation is unique to the medical setting; b) that the stress potential of a problem and the need for a professional solution affect independently the propensity of assessing the responses in such a manner; c) that the judgment of the mode indeed determines the assessment of the quality of the content of the support and is not simply correlated with it; and d) that this pattern of assessment is not confined, as indicated in the previous model,[7,8] to the patient–physician interaction, but may apply also to the interaction with other roles in the medical setting. We believe that the present study may be considered as an additional contribution to the understanding of the conditions required for achieving satisfaction with the treatment, which in turn may be conducive to the achievement of the goals of the treatment.

References

1. Korsch BM, Freeman B, Negrete VF. Practical implications of doctor-patient interaction: analysis for pediatric practice. AJD. Ch: 1971;121:110.

2. Janis IL. Stress and Frustration. New York: Harcourt, Brace and Jovanovich, 1971.

3. Ley P, Spelman MS. Communicating with the Patient. London: Staples, 1967.

4. Egerg J, Battit GE, Welch CE, Battlett MK. Reduction of postoperative pain by encouragement and instruction of patients. N Engl J Med 1964:270:825.

5. Skipper J, Leonard R. Children stress and hospitalization: field experiment Journal of Health and Human Behavior 1968;9:275.

6. Spiro A. Factors contributing to the placebo effect. A J Psychother 1964;43:73.

7. Ben-Sira Z. The function of the professional's affective behavior in client satisfaction: a revised approach to social interaction theory. J Health Soc Behav 1976;17:3.

8. Ben-Sira Z. Affective and instrumental components in the physician–patient relationship: an additional dimension of interaction theory. J Health Soc Behav 1980;21:170.

9. Beecher H. The powerful placebo. JAMA 1959;159:602.

10. Cartwright A. What goes on in the general practitioner's surgery? Acheson RM, Aird L, eds. Seminars in Community Medicine, Vol. 1: Sociology. London: Oxford University Press, 1976;23-48.

11. Kasl S, Cobb S. Health behavior, illness behavior and sick-role behavior. Arch Environ Health 1966;12:246, 531.

12. Kosa J, Robertson LS. The social aspects of health and illness. In: Kosa J, Antonovsky A, Zola IK, eds. Povery and health: A sociological analysis. Cambridge, MA: Harvard University Press, 1969;35-68.

13. Rosenstock IM. Why people use health services. Milbank Mem Fund Q 1966;44:Part 2.

14. Duff R, Hollingshead A. Sickness and Society. New York: Harper and Row, 1967.

15, Levine G. Anxiety about illness: psychological and social bases. Journal of Health and Human Behavior 1962;3:30.

16. Lewis L, Lopreato J. Arationality, ignorance and perceived danger im medical practices. American Sociological Review 1962;27:508.

17. Balint M. The Doctor, his patient and illness. London: Tavistock, 1957.

18. Johnson TJ. Professions and power. London: Macmillan, 1972.

19. Weiss RS, Harwood E, Riesman. The world of work. In: Merton RK, Nisbet R, eds. Contemporary social problems. 4th ed. New York: Harcourt, Brace, Jovanovitz, 1972;603-37.

20. Dohrenwend BS, Dohrenwend BP. Class and race as status related sources of stress. In: Levine S, Scotch NA, eds. Social Stress. Chicago: Aldine, 1970;111-40.

21. Antonovsky A. Health, stress and coping. SanFranciso: Jossey Bass, 1979.

22. Gross E. Work organizations and stress. In: Levine S, Scotch NA, eds. Social stress. Chicago: Aldine, 1979; 54-110.

23. Rave A. Finding periodic patterns in time series with monotonic trend: a new technique. In: Shye S, ed. Theory construction and data analysis in behavioral sciences. San Francisco: Jossey-Bass, 1978;371-90.

24. Guttman L, Levy S. Structure and dynamics of worries. Sociometry 1975;38:448.

25. Guttman L, Levy S. Worry, fear and concern differentiated. Israel Annals of Psychiatry and Related Disciplines 1976;14:211.

Patient Suffering and Patient Satisfaction Among the Chronically Ill

LAWRENCE S. LINN, PH.D., AND SHELDON GREENFIELD, M.D.

In a patient population with serious, debilitating chronic disease, the present study tested the hypothesis that a significant amount of variance in patient satisfaction ratings of provider behavior would be explained by patients' current assessment of their health and their social circumstances. More favorable patient ratings of the art, technical quality and efficacy of their care givers were found to be significantly associated with more positive general health perceptions, fewer days spent in bed due to health problems and lower scores on a widely used depression scale, as well as with age and ethnic background. These data provide evidence indicating that patient satisfaction measures are sensitive to and confounded by patients' perceived health, view of life and social circumstances. It is suggested that if patient satisfaction ratings are to be used as indicators of the quality of health care delivery, there is a need to account for the variance attributable to measurable psychological, social and demographic factors.

MEASURING PATIENT SATISFACTION has become an extremely popular way of assessing health services. It is generally assumed that patient ratings and evaluations of provider behaviors and services are a direct and accurate reflection of health care delivery.[1,2] However, one study has found that patients who were less satisfied with their life in general, or with the community facilities and resources, were also less satisfied with the medical care they were receiving.[3] This study led us to consider whether patient dissatisfaction may be in part related to other measurable factors, particularly the current affective orientation of the patient.

An opportunity to test the hypothesis that the ratings and evaluations of provider behavior are related to the patient's current perception of his/her social, physical and mental health arose in connection with a study of three subspecialty clinics and a large general ambulatory care center at a major teaching hospital. The extreme amount of discomfort reported and observed among chronically disabled outpatients who seek care in a university-based medical center[4-7] provided an excellent opportunity to explore how patient attitudes and ratings of their own health relate to the way they evaluate their health care providers.

Research for this article was supported in part by a grant (#5785) from the Robert Wood Johnson Foundation.

Address for reprints: Dr. Lawrence S. Linn, Department of Medicine, UCLA, Center for the Health Sciences, Los angeles, CA 90024.

From the Department of Medicine and School of Public Health, UCLA, Los Angeles, California.

0025-7079/82/0400/0425/$00.80 © J. B. Lippincott Co.

Method

Sample

The patients in the present study were sampled from three medical subspecialty clinics (Arthritis, Cardiology and Endocrinology), and a large general medical clinic (Medical Ambulatory Care Center) at the University of California in Los Angeles. For the subspecialty clinics, the sample consisted of all English-speaking patients who made at least one follow-up visit to one of the clinics between March 1, 1980 and October 31, 1980. Names were drawn from the appointment logs of each clinic, but new patients, cancellations, and people who did not show up were excluded.

For the Medical Ambulatory Care Center, the patient sample was also drawn from appointment logs, but sampling was based upon a random selection of clinic-days during the study period. Once again, all patients who kept a follow-up appointment on the selected clinic sessions became part of the sample population. Therefore, all patients in the sample are considered regular or continuing care patients.

In total, 804 patients were sampled and received questionnaires by mail (Arthritis, N = 186; Cardiology, N = 142; Endocrinology, N = 143; and General Medical, N = 333). After follow-up telephone calls to nonrespondents and a second mailing of the survey to them, completed and usable questionnaires were returned by 519 patients, constituting a 65 per cent response rate.

Design

The charts of all patients in the subspecialty clinic samples were examined in order to ascertain the patients' most current phone numbers and addresses. Most phone numbes and addresses for Medical Ambulatory Care patients were available on the clinic appointment logs. All patients eligible to be in the study were telephoned by a research assistant who explained the purpose of the study and informed them about the consent process required by the University. A script was carefully followed by the research assistant as a guide for answering patient questions about the study. If a patient could not be reached after three attempts, the questionnaire was sent to the patient. For patients who did not complete and return questionnaires within three weeks, follow-up phone calls that were also guided by a script calls were made.

Measurements

Patient Satisfaction. Three dimensions of satisfaction based upon the work of Ware et al.[1] and used by us in a previous study[8] were measured: 1) *Art of Care*, a nine-item scale assessing provider characteristics and behaviors such as interest, friendliness, courtesy, regard for privacy and degree of caring (Alpha reliability = 0.84); 2) *Technical Quality*, a six-item scale focusing on the competence of physicians in terms of their knowledge, thoroughness, skill and experience (Alpha reliability = 0.86); and 3) *Efficacy*, a four-item scale indicating the degree to which patients felt that doctors helped them to get well or feel better (Alpha reliability = 0.83). Each of these scales had positively and negatively worded statements scored on a five-point scale from strongly agree to strongly disagree. (The specific items are shown in the Appendix.)

Health Status. Five different measures were used: 1) *Disability Days*, a single question asking patients to report the number of days in the last month they had to stay in bed most or all of the time due to their health; 2) *General Health*, a four-item scale composed of ratings of pain and worry related to health and ratings of their general overall health (Alpha reliability = 0.84); 3) *Physical Abilities and Limitations*, a scale based on reports of patients' ability to perform nine activities such as walking, climbing stairs, working or driving a car (Alpha reliability = 0.85); 4) *Social Health*, a four-item scale assessing the frequency of social contact with close friends or family over the past month (Alpha reliability = 0.75); and 5) the *Zung Self-Rating Depression scale*, a 20-item mental health measure of depression.[9,10] The previous four measures of health status were taken from the Rand Corporation Health Insurance study.[11,12]

Situational Variables. In addition to measures of perceived and reported health, the present study also explored the relationship between patient satisfaction and seven situational or demographic characteristics: age, sex, educational background, racial/ethnic background, marital status, current employment status, and the mode of transportation taken to and from the medical center.

Results

Prevalence of Satisfaction

Although a range of satisfaction was expressed by the patients sampled, ratings of care were high. All three scales had mean item scores around 4.0 on a five-point scale. Greatest satisfaction was with the Art of Care (mean 4.08); least satisfaction was expressed on the Efficacy scale (mean 3.90). Intercorrelations among the three satisfaction scales were high and statistically significant, ranging from 0.75 to 0.79.

Prevalence of Suffering

The patients in the present study reported a considerable amount of pain, suffering and disability. For example, 16 per cent reported having been in bed most or all of a day in the previous month. On the General Health scale, only 6 per cent rated their health as being "excellent" and 31 per cent as "good," the vast majority (63 per cent) rated their health as being "fair" or "poor." Also, 56 per cent reported having been in some or a great deal of pain in the past three months, and 90 per cent indicated that their health had worried or concerned them during that period.

Based on responses from the Physical Abilities and Limitations scale, 65 per cent of the patients had trouble walking several blocks or climbing stairs because of their health, 62 per cent had trouble working, and 13 per cent could not drive a car for health-related reasons. On the Social Health scale, 34 per cent of the patients were not visited by a friend in the past month, and 34 per cent did not themselves visit a friend. Finally, on the Zung Self-Rating Depression scale, only 45 per cent of the patients had scores in the normal range. Twenty-six per cent had scores indicating mild depression, 20 per cent moderate depression and 8 per cent were severely depressed.

Table 1 shows the intercorrelations among the health status indicators. Although all five indices are significantly correlated with each other, the four-item General Health scale had the strongest association with the other health measures. The Social Health scale had the weakest correlations with other indicators.

Situational Factors and Satisfaction

There were no significant associations between the three satisfaction measures and either sex or the mode of transportation used to get to the medical center. However, age was significantly correlated with all satisfaction measures ($p < 0.001$). As age increased, ratings of satisfaction increased. Educational background was not related to satisfaction with the art or technical quality of care. However, patients with less education were significantly more satisfied with the efficacy of care ($p < 0.05$).

One-way analysis of variance indicated that ethnic background was significantly

TABLE 1. Intercorrelations Among Health Status Indicators: Disability Days (DD), General Health (GH), Physical Abilities and Limitations (PAL), Social Health (SH) and Depression (D)

	DD	GH	PAL	SH	D
Disability days	—	−0.525	0.407	−0.129*	0.381
General health		—	−0.710	0.251	−0.585
Physical abilities and limitations			—	−0.269	0.499
Social health				—	−0.295
Depression					—

* $p < 0.01$; all other correlations $p < 0.0001$.

associated with satisfaction with the art of care ($p < 0.001$), the technical quality of care ($p < 0.09$) and the efficacy of care ($p < 0.01$). Consistently, black and Hispanic patients were more satisfied with their health services than Caucasian patients. With regard to marital stutus, single patients and those separated or divorced were the least satisfied with the technical quality ($p < 0.05$) and efficacy of care ($p < 0.01$). Widowed patients, and to a lesser extent those who were married, were the most satisfied on these two dimensions of care. Art-of-care ratings were not related to marital status, although the trends reflected the patterns of association of the other satisfaction measures.

Finally, employment status was found to be significantly associated with all satisfaction measures ($p < 0.05$ or $p < 0.01$). Retired patients were consistently the most satisfied; those unemployed were the least. Patients who worked expressed levels of satisfaction between these two groups. However, there were no apparent differences between full-time and part-time workers.

Suffering and Satisfaction

Table 2 shows the correlations between measures of health status and ratings of satisfaction. Patients who reported having spent more days in bed during the past month because of their health were consistently less satisfied with the health care they had been receiving on all three scales ($p < 0.0001$). Similarly, patients who were more depressed and who rated their general health as poorer were significantly

TABLE 2. Pearson Correlation Coefficients Between Health Status Measures and Ratings of Patient Satisfaction (N = 500)

	Satisfaction Ratings		
Health Status	Art of Care	Technical Quality	Efficacy
Number of days disabled	−0.181 *	−0.242 *	−0.245 *
General health	0.222 *	0.266 *	0.324 *
0.2222⅛	0.266*	0.324*	
Physical abilities/limitations	−0.065	−0.107†	−0.135†
Social health	0.044	−0.006	0.022
Depression (Zung SDS)	−0.215*	−0.209*	−0.263*

* $p < 0.0001$.
† $p < 0.01$.

less satisfied with their care ($p < 0.0001$). Patients with more physical limitations were also significantly more likely to express dissatisfaction with the technical quality and the efficacy of their care ($p < 0.01$). However, these correlations were much lower than those of the other health status indicators, and satisfaction with the art of care was unrelated to physical ability. Lastly, none of the satisfaction ratings was significantly associated with the social health measure.

Because of these significant associations and those reported between situational and demographic factors in the last section, it was decided to perform a step-wise multiple regression analysis for each of the satisfaction measures. As seen in Table 3, the General Health scale and the ethnicity of patients consistently and independently explained a significant amount of variance in all three satisfaction scales. Age also explained a significant amount of variance in two of the three measures. In summary, about 12 per cent of the variance in art-of-care ratings was explained by health status and situational variables, about 13 per cent of the variance in ratings of technical quality and 21 per cent of the variance in the efficacy ratings.

TABLE 3. Step-wise Multiple Regression Analysis for Variables Significantly Associated With Patient Satisfaction Ratings

Variables Entered	Multiple R	R^2
Art of Care		
General health	0.2449	0.0600
Ethnicity *	0.3181	0.1012
Age	0.3412	0.1164
Technical Quality		
General health	0.2983	0.0890
Ethnicity *	0.3286	0.1080
Marital status: being single	0.3533	0.1248
Disability days	0.3658	0.1338
Efficacy		
General health	0.3385	0.1146
Ethnicity *	0.3965	0.1572
Age	0.4377	0.1916
Driven to medical center	0.4536	0.2057
Depression	0.4621	0.2135

* Patients classified as racial minority or not.

Discussion

The present study was not designed to be a comprehensive or exhaustive analysis of factors associated with patient satisfaction. Rather, the purpose was to determine the degree to which evaluations of health care providers were correlated with situational characteristics and perceptions of health status among a patient population in which pain, suffering and disability were highly prevalent. The present study has documented that ratings of health were significantly associated with patient evaluations of their regular providers of health care services. Furthermore, the four-item measure of General Health explained most of the variance in such satisfaction ratings, more than longer physical and mental health measures. This finding, together with those linking satisfaction ratings to ethnicity, age and marital status, raises some concern about the validity of patient satisfaction ratings as a measure of the quality of health care delivery. We conclude that when patient ratings of physician behavior are to be used as indicators of the quality of care, the feelings of patients about their health, as well as characteristics of their social situation, need to be included as control variables.

References

1. Ware JE Jr, Davies-Avery A, Stewart AL. The measurement and meaning of patient satisfaction. Health and Medical Care Services Review 1978;1:2.

2. Spitzer WO, et al. Evaluating new methods for provision of primary care: an Ontario strategy. Med Care 1978;16:560.

3. Linn LS. Factors associated with patient evaluation of health care. Milbank Mem Fund Q 1975;53:531.

4. Fletcher SW, Fletcher RH, Pappius EM, Rudd R.A teaching hospital medical clinic: secondary rather than primary care. J Med Educ 1979;54:384.

5. Rudd P, Tul V, Brown K, et al. A general medicine clinic: the dilemma and teaching implications. J Med Educ 1979;54:766.

6. Linn LS, Yager J. The effect of screening sensitization and feedback on notation of depression. J Med Educ 1980;55:942.

7. Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. J Fam Prac 1978;7:509.

8. Linn LS, Ware JE, Greenfield S. Factors associated with relief from chest pain following emergency care. Med Care 1980;18:624.

9. Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965;12:63.

10. Hedlund, JL, Vieweg, BW. The Zung self-rating depression scale: a comprehensive review. J Oper Psychiatry 1979;10(No. 1):51.

11. Brook RH, et al. Conceptualization and Measurement of Health for Adults in the Health Insurance Study, Vol. VIII, Overview. Santa Monica: Rand corporation, 1979. (#R-1987/8-HEW.)

12. Ware JE Jr, Davies-Avery A, Brook RH. Conceptualization and Measurement of Health for Adults in the Health Insurance Study, Vol. VI, Analysis of Relationships Among Health Status Measures. Santa Monica: Rand Corporation, 1980. (#R-1987/6-HEW.)

Appendix

Frequency distributions of responses to items measuring three dimensions of patient satisfaction: art of care, technical quality and efficacy of care.*

	Per Cent Agree	Per Cent Uncertain	Per Cent Disagree
Art of Care			
The doctors seemed to have a genuine interest in me as a person.	82	10	9
The doctors did their very best to keep me from worrying about things.	81	10	9
The doctors were very friendly and seemed devoted to me and other patients.	88	7	5
In general, all the clerks and receptionists were courteous and respectful.	87	5	8
No one seemed to care very much about my privacy.	6	9	86
The nurse treated me in a friendly manner.	95	3	2
The doctors took time to put me at ease and tried to make me more comfortable.	89	6	6

* Categories of strongly agree and agree were combined, as were strongly disagree and disagree.

The doctors caused me to worry because they did *not* explain things to me or tell me what was happening.	11	6	83
Doctors in this clinic don't want to bother with many of the things that really trouble me.	13	11	76
Technical Quality			
The doctors were not as thorough as they should have been.	16	10	73
The doctors seemed to lack experience with my kind of problem.	16	13	70
The doctors seemed very organized (knew what to do next and had my treatment well planned).	71	15	14
The doctors were very careful to check everything when examining me.	82	8	10
The quality of the care I received there was the best anyone could expect.	81	12	8
The doctors seemed very competent and well trained.	86	11	3
Efficacy			
The doctors helped me feel a lot better.	76	10	14
I am in better health now because of the care I received there.	70	17	12
The doctors did as much as could be expected to help me get well.	84	11	6
Some of the things the doctors did were *not* very helpful.	18	15	67

MEDICAL CARE
April 1982, Vol. XX, No. 4

Communication

Do Patients Want Access to Their Medical Records?

MAX MICHAEL, M.D., AND CLAY BORDLEY, B.S.

Patient access to the medical record has a demonstrated effect on patient satisfaction and the physician–patient relationship. A study of 200 patients attending an outpatient clinic confirmed that most patients would like to receive a copy of their record, while slightly more than half felt there was a relationship between having the record and quality of care. The implications of giving patients a copy of their record is discussed.

THE DEBATE ABOUT PATIENT ACCESS to the medical record was focused sharply in 1973 when Shenkin and Warner proposed legislation requiring that a "complete and unexpurgated copy of all medical records, both inpatient and outpatient, be issued routinely and automatically to patients as soon as the services provided are recorded."[1] The potential advantages and disadvantages of any system that encourages patient access to the medical record are unsolved despite federal legislation,[2] the occasional implementation in a practice setting,[3,4] and the American Hospital Association's Patient's Bill of Rights.[5] Evidence that would suggest improvement in care when patients receive a copy of their medical record is meager.[4,6,7] Moreover, it is universally assumed that patients would indeed like to see their records and that such disclosure would benefit their medical care, and by implication, their health.

Unfortunately there has been no systematic study to ascertain patients' a priori feelings about receiving their medical records. The present study was undertaken to ask patients if they actually wanted to see their records and if they felt that reading their records would improve the quality of the care they receive from their physician. The hypothesis tested was that in the setting described patients would want to read their medical records but not necessarily to obtain an updated copy.

Methods

Two hundred patients attending the outpatient clinics of a municipal hospital of moderate size were given a short questionnaire by one of us (C.B.) during the month of May 1980. These patients were randomly selected from approximately 4,400 patients seen during that month. Each clinic area (Medicine, Surgery, OB–GYN, Ophthalmology, ENT and Dermatology) and each meeting time (day of week, and morning or afternoon) were equally represented in the sample.

The questionnaire contained four simple questions pertaining to the medical record. Revision through pretest analysis

---

From the Cooper Green Hospital, Birmingham, Alabama.

Request for reprints: Max Michael, M.D., Cooper Green Hospital, 1515 6th Avenue South, Birmingham, AL 35233.

0025-7079/82/0400/0432/$00.70 © J. B. Lippincott Co.

of the questions was accomplished three times.

Prior to individual administration of the questionnaire, a standard, brief introduction explained as clearly as possible the purpose of the questionnaire and attempted to make certain that the patient knew what was meant by the medical record. No one refused to complete the questionnaire; all were usable.

Results

The patients surveyed were identical to those typically served by the clinics. Medicare or Medicaid recipients accounted for 48 per cent of the patients seen, 32 per cent are county-supported indigent persons without any form of third-party coverage, 16 per cent earn salaries above the poverty level, and approximately 4 per cent are vocational rehabilitation clients, inpatients and the like. Of the patients surveyed, 74 per cent were black, 26 per cent were white, 71 per cent were females, and 29 per cent were males; the average age was 47.

Table 1 displays the responses to the four questions; each question is described in detail below.

Question 1. Should Patients Be Permitted to See Their Own Medical Records?

Almost 80 per cent of the patients responded affirmatively to this question. The responses did not vary significantly among men and women or age groups. The apparent difference between black and white respondents was not significant.

Question 2. Would Reading Your Medical Record Help You Understand What Is the Matter With You and What the Doctor Is Doing About It?

Sixty-seven per cent of the respondents felt that reading the chart would improve their understanding of their medical problems. There was not a significant variation in response among women and men or age and racial group.

Question 3. Would Reading Your Medical Record Improve the Quality of Medical Care You Receive?

Slightly more than half (54 per cent) of the patients felt that there was a positive relation between reading their chart and quality of care as they perceived it. Again, no difference in sex, age or race was evident in the responses.

TABLE 1.

Question		Response (%)*			
		Female	Male	White	Black
1. Should you see your record?	Yes	74	84	89	73
	No	23	12	11	23
	No opinion	3	4	0	4
2. Would reading it help you understand your illness?	Yes	65	72	64	69
	No	30	16	28	25
	No opinion	5	12	8	6
3. Would reading it improve quality of care?	Yes	54	53	55	53
	No	41	36	43	37
	No opinion	5	11	2	10
4. Should you have a personal copy?	Yes	52	48	36	57
	No	44	45	62	37
	No opinion	4	7	2	6

* Some columns add to more than 100 per cent due to rounding. Differences between all groups are insignificant at $p > 0.05$.

Question 4. Should You Be Given A Copy of Your Medical Record to Take Home With You if You Want It?

Fifty-two per cent agreed that a personal copy of the medical record should be made available if desired. There was no response difference between men and women, and although 36 per cent of white respondents, compared with 57 per cent of black respondents, answered yes to the question, the difference did not achieve statistical significance. No differences were apparent by sex within these racial groups or by ages.

Discussion

From the results of this study it is clear that patients generally desire access to their medical record and that more than half would want a personal copy. In addition to the questionnaire data, comments made by a number of patients suggested that they were not convinced that they could keep up with a copy of the chart or that having a copy was necessarily as important as simply reading it.

With this in mind, it has not been clearly demonstrated that patients should be given a copy of their medical records to take home in addition to an opportunity to see it in the doctor's office or at the hospital.

Although some medical practices routinely give patients a copy of their record, few studies have examined the impact of this practice. In a study of 125 patients admitted to a 16-bed rehabilitation medicine service, Golodetz, Ruess and Milhous[6] showed that the majority of the patients preferred having a copy of their record rather than a formal discussion with their doctor. Many made additions or corrections (50 per cent) or asked questions on vocabulary or meaning (60 per cent). There were no instances noted of patients unable to handle the information contained in the record.

Stevens and associates,[8] in a study of 25 hospital patients who were given access to their medical records, found that "comparisons of attitudes toward doctors, nurses, hospitals, medications, illness, and self showed no significant differences between the open-record group and a similar sized control group . . ." Only seven of the patients requested to see their records at times other than during the formal rounds, 17 received them passively during rounds, and one refused to look at her record altogether, insisting that she be told everything by her doctor. The authors felt that perhaps the patients did not feel that reading their records was going to improve either the care they received or their health, or they were not really concerned with the information contained in the records.

A study of patients who requested to see their medical record likewise showed little impact on the care received.[9] Only 11 patients from approximately 2,500 admissions during the time under study requested to see their charts. In no instance did receipt of the chart improve the quality of care or lead to the deterioration of the physician–patient relationship.

Finally, at the Given Health Care Center in Vermont, patients routinely receive a copy of clinic notes. Bouchard and Associates[3] have shown high levels of patient acceptance, as well as an apparent reduction in anxiety about health-related matters.

The present study is a first attempt at asking patients before they are actually given a copy of the record whether they think they should be allowed to read their medical records and whether they think such access would improve their understanding of their illness and its treatment. It has been found that patients are in fact interested in seeing their records "if they want to." However, the real question, whether or not they would want to see the records, remains.

In the preface to the questionnaire, patients were asked to assume that they could read their records and would have explained to them anything they did not understand. For this to be possible, several obvious changes would have to take place in today's record-keeping methods, all of which would cost both time and money (two particularly tight resources in the clinical setting). First, physicians would have to take more care in their note taking or have dictated notes. Second, records would have to be copied or reduced, and perhaps paraphrased. These processes would obviously add to the already high costs of record keeping. In our clinical setting, at least, the ability of the patient to understand much of what goes into the record is limited, thereby limiting the usefulness or benefit of record sharing. Furthermore, what appears in the medical record is usually an inadequate indicator of the events occurring in the visit.

Despite these obvious drawbacks, the apparent advantages of record sharing are impressive. Physicians and patients seem to get along better, and patients are often quite pleased with the information and attention they receive.[10] We cannot assume, however, that patients are longing to see their records and that granting them that access is necessarily good. Although the notion of showing the record has intuitive appeal, the dangers of implementing a procedure, test or treatment without proper evaluation under reasonably controlled circumstances are well known. Is there in fact any relationship between record sharing and improved patient health, other than patient satisfaction? The answer to this question may come from an ongoing, longitudinal prospective study.[11]

Satisfaction is necessary to the medical encounter, but it remains conjecture that access to the medical record is an important aspect of improving satisfaction, the physician–patient relationship or quality of care. Improved satisfaction is attributable equally to the Hawthorne effect and to possession of the medical record.

Herein must lie the central concern for those interested in the issue of patient access to the medical record. Until a connection between an important health-related aspect of medical care and the medical record is demonstrated, it is unreasonable to toss the record into the patient's lap in order either to improve understanding or to stave off apprehensions. The solution to problems inherent in the present day physician–patient interface might not be found through record sharing. Before the physician-related aspects of record keeping are diluted or lost, we must ask who should get a copy? Why should they get a copy? What are the expected benefits? And what are the potential harms? These questions deserve an answer before any health system adopts an intuitively pleasing idea.

References

1. Shenkin BN, Warner DC. Giving the patient his medical record: a proposal to improve the system. N Engl J Med 1973;289:688.

2. The Federal Privacy Act of 1974. Public Law 93-579.

3. Bouchard RB, Tufo HM, et al. The patient and his problem-oriented record. In: Walker HK, Hurst JW, Woody MF, eds. Applying the problem-oriented system. New York: Medcom Press, 1973:42.

4. Hertz CG, Bernheim JW, Perloff TN. Patient participation in the problem oriented system: a health care plan. Med Care 1976;14:77.

5. A Patient's Bill of Rights, American Hospital Association, 1972.

6. Golodetz A, Ruess J, Milhous RL. The right to know: giving the patient his medical record. Arch Phys Med Rehabil 1976;57:78.

7. Sherman FT, Libow LS. A portable medical record system for the elderly. JAMA 1979;242:57.

8. Stevens DP, Staff R, Mackay LR. What happens when hospitalized patients see their own records. Ann Intern Med 1977;86:474.

9. Altman JH, Reich P, Kelly MS, et al. Patients who read their hospital charts. N Engl J Med 1980;302:169.

10. Westin AF. Medical records: should patients have access? Hastings Cent Rep December 1977:23.

11. Shapiro M, Lehman A. Patients who read their hospital charts (letter). N Engl J Med 1980;302:1482.

C5170-1 5/21/82

Medical Care

May 1982
Vol. 20 No. 5
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

A Current Reappraisal of Berry's Hospital Typology

Improved Cost Allocation in Case-Mix Accounting

Three Case-Type Classifications: Suitability for Use in Reimbursing Hospitals

A Patient-Classification System for Long-Term Care

Comparing Classification Methods: Measurement of Variations in Charges, Length of Stay, and Mortality

The Measurement of Hospital Case Mix

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

MEDICAL CARE
May 1982, Vol. XX, No. 5

Improved Cost Allocation in Case-Mix Accounting

SANKEY V. WILLIAMS, M.D.,[*][†] STEVEN A. FINKLER, PH.D., C.P.A.[†][‡]
CATHERINE M. MURPHY, M.H.A.,[§] AND JOHN M. EISENBERG, M.D., M.B.A.[†][∥]

Traditionally, many hospital costs have been allocated to patients using indirect measures that do not always reflect the value of the resources used to provide care. When, for example, costs are allocated by multiplying the patient's charges by the hospital's ratio of costs to charges, the allocated cost does not reflect actual cost because the hospital does not uniformly charge for services in proportion to their cost. The choice of method for cost allocation will be as important for the newly developed case-mix cost-accounting systems as it has been for traditional cost-accounting systems. To illustrate how the use of an indirect cost-allocation method might affect the output of a case-mix cost-accounting system, operating room, radiology and clinical laboratory costs were assigned to 106 hospitalized inguinal hernia patients in two diagnosis-related groups (DRGs) using both the hospital's existing cost-allocation method and a method that measures costs directly. Total costs and the costs in each department were significantly lower in each DRG using the direct method. It was concluded that patients in these two DRGs were being assigned more than the actual cost of their care with the existing cost-allocation method and, therefore, that the existing method prevented the case-mix accounting system from providing accurate management information.

RECENT INTEREST in the control of hospital costs has led to increased concern about how the costs of medical services are determined. In the past, attempts to associate costs with the hospital's products have not been convincing because hospital products are difficult to define. Simple measures of productivity, such as the number of patients admitted or the number of patient days, remain unsatisfactory because of the wide variation in inputs and costs from patient to patient.[1-4] To address the problem of defining hospital products, a patient classification system that separates the patients into 383 diagnosis-related groups (DRGs) has been proposed.[5-10] Each DRG is composed of a group of patients with similar diagnoses and lengths of stay. The hospital's products are defined as the bundle of services that the hospital provides to each of the 383 groups, and specific costs are assigned to each bundle.

The DRG system is only one of several case-mix measures that have been pro-

---

[*] Henry J. Kaiser Family Foundation Faculty Scholar, Section of General Medicine; and Associate Director for Medical Affairs, National Health Care Management Center, University of Pennsylvania, Philadelphia, Pennsylvania.

[†] Senior Fellow, Leonard Davis Institute of Health Economics.

[‡] Assistant Professor of Accounting, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania.

[§] Associate Administrator, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania.

[∥] Sol Katz Associate Professor of Medicine and Chief, Section of General Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Supported by the National Health Care Management Center at the Leonard Davis Institute of Health Economics, University of Pennsylvania, through Grant No. HS-02577 from the National Center for Health Services Research, Office of the Assistant Secretary of Health, Department of Health and Human Services.

The opinions, conclusions and proposals in this paper are the responsibility of the authors, and do not necessarily represent those of the organizations that supported this work.

Reprint requests: Sankey V. Williams, M.D., Section of General Medicine, Box 624, Hospital of the University of Pennsylvania, Philadelphia, PA 19104.

0025-7079/82/0500/0450/$01.00 © J. B. Lippincott Co.

posed, but it is the most developed and most widely used.[11] Significant problems are being identified with the DRG system as more experience is gained using it,[11-14] especially as a method for reimbursing hospitals prospectively for patient care.[10,15-17] Nevertheless, case-mix measures such as the DRG system probably represent a significant step forward for hospital management because they permit the hospital administrator to control operations, to plan and to budget more effectively, and they may facilitate the planning of regional services.[9,10,18] However, the new method's potential cannot be realized if costs are not accurately assigned to each DRG.

At present, hospitals generally assign most patient-care costs to functional cost centers, such as the departments of radiology, nursing and maintenance. Costs are then assigned to patients based on their use of each department's services or on a per diem basis. Assigning costs on a per diem basis means that all patients are assigned an average cost regardless of the cost of the services actually used. Moreover, even when patients are assigned costs based on their use of services, the cost-allocation method often does not recognize the real differences in cost among different services, but substitutes costs that have been measured indirectly and thus may be inaccurate. Both of these approaches obscure the relationship between a group of patients and the resources they consume. Unfortunately, these same cost-allocation methods have been used to assign costs to the products defined by the DRG system.[8,10,16]

Allocation of operating room costs illustrates the problem. Consider how surgery charges are usually determined. All operating room costs are accumulated, and the total is divided by the total minutes of operating time. The charge is then calculated, based on a cost per minute of time in the operating room. Operating time, however, is an inadequate measure for resource consumption. Some short procedures require more expensive equipment, more supplies and more total personnel time than do longer procedures. This problem might be avoided if more direct measures were used to assign operating room costs.

To illustrate how this might change the allocation of costs, we report here a pilot project using a cost-allocation method that assigns many departmental costs directly to patients based on cause and effect. The new system is compared with a traditional allocation system that differs from the new system in that the traditional system assigns departmental costs to patients indirectly, based on the departmental cost-to-charge ratio multiplied by the patients' departmental charges.[19]

Methods

The project was conducted at the Hospital of the University of Pennsylvania, a 694-bed university teaching hospital. Patients in two diagnostic-related groups were included, DRG 198 and DRG 199. Patients in both groups had a primary discharge diagnosis of inguinal hernia, without obstruction, and were between the ages of 14 and 65 years. Patients in DRG 198 did not and patients in DRG 199 did have a major secondary diagnosis.

The medical records of all patients discharged in 1978 with a primary diagnosis of inguinal hernia were reviewed. Of the 192 patients whose discharge abstracts were coded for inguinal hernia, 60 were over the age of 65 and thus were not included in the two DRGs by definition, 15 actually had a diagnosis other than inguinal hernia, and for 11 it was not possible to match the record with the patient's bill. Of the remaining 106 patients included in the study, 69 were in DRG 198 and 37 were in DRG 199. All 106 patients underwent operations to repair their inguinal hernias.

Two methods were used to allocate costs to the 106 patients. Both cost-allocation methods were restricted to three cost centers—the operating room, radiology and the clinical laboratories. Other cost centers were not examined, and hospital overhead costs, such as hospital administration and building depreciation, were excluded from the calculations.

The Direct Method

The direct cost-finding method is a variation of the component-enumeration method,[4] which was adapted to this project.

With the direct method to allocate costs to each patient, the first step was to divide the costs from each functional cost center into 1) departmental overhead, such as office supplies, equipment depreciation, and salaries for departmental administrators and supervisors; and 2) direct costs, such as supplies and salaries, that were committed to direct patient care (Fig. 1).

The second step was to choose a way of allocating the departmental overhead to individual patients. For the operating room, the departmental overhead was divided by the total number of operations to yield an overhead cost per operation, which was assigned to each patient. For the radiology department and the clinical laboratories, departmental overhead was divided by the number of procedures, and each patient was assigned a departmental overhead cost that was proportionate to the number of procedures performed in the inguinal hernia patients.

The third step was to assign direct departmental supplies and salaries to inguinal hernia patients. For the operating room, a list of supplies consumed during a typical inguinal hernia operation was prepared and verified by interviewing operating room nurses, supply clerks and surgeons who performed inguinal hernia operations. These supplies were designated as either variable or fixed supplies. The variable supplies consisted of disposable items, such as medical–surgical supplies and paper sheets and gowns that were used only once. They were counted, and their costs were determined for assignment to each patient.

It was more difficult to assign the cost of fixed supplies to each patient. The fixed

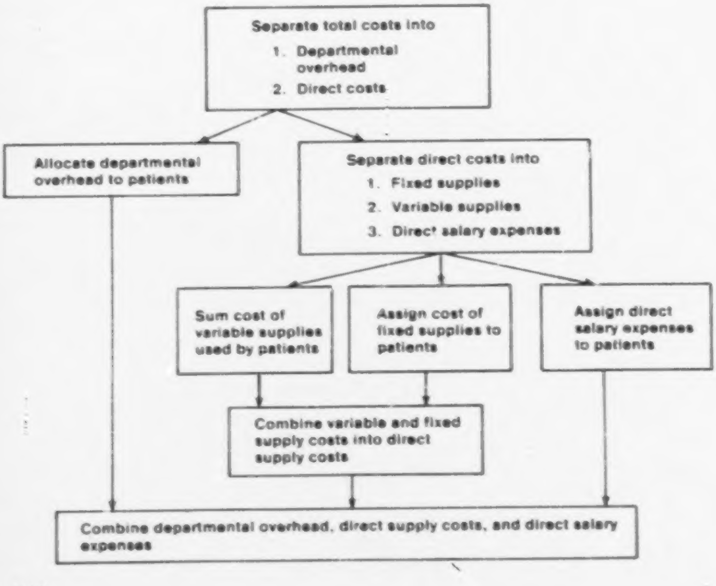

Separate total costs into
1. Departmental overhead
2. Direct costs
Allocate departmental overhead to patients
Separate direct costs into
1. Fixed supplies
2. Variable supplies
3. Direct salary expenses
Sum cost of variable supplies used by patients
Assign cost of fixed supplies to patients
Assign direct salary expenses to patients
Combine variable and fixed supply costs into direct supply costs
Combine departmental overhead, direct supply costs, and direct salary expenses

FIG. 1. The direct method for cost allocation.

supplies consisted of laparotomy instrument sets, major large basin sets and other reusable supplies. The cost per operation of fixed supplies, such as instrument sets, was determined by multiplying the cost of each instrument set by the number of sets to get a total cost, and then dividing by the number of operations in which those supplies were used and correcting for the expected useful life of the set in years. Variable and fixed supply costs were added together to yield the total direct supply cost, which was then assigned to each patient.

To assign direct salary expenses in the operating room, interviews and observation were used to determine how many hospital employees worked with a typical inguinal hernia patient. An average of one circulating and one scrub nurse worked with each patient throughout the operation. Anesthesia time was obtained from the anesthesia records in each medical record. The average operating room nurse salary for 1978 was inflated by 30 per cent to account for fringe benefits. The hourly salary for two operating room nurses was multiplied by total anesthesia time, resulting in total departmental salary expense per operation.

The costs of direct departmental supplies and salaries in radiology and the clinical laboratories were assigned to inguinal hernia patients using existing measures. The radiology department has determined the direct costs for x-ray procedures based on measures of personnel time and the consumption of supplies similar to those used for the operating room. In the clinical laboratories, measures of direct costs that were based on the amount of personnel time required to complete each test as published by the College of American Pathologists were used.[20] In both departments, the direct cost of each procedure was multiplied by the recorded number of procedures used by inguinal hernia patients to arrive at the total costs of direct departmental supplies and salaries.

To summarize the direct method, the total cost allocated to inguinal hernia patients included the costs of 1) departmental overhead; 2) direct departmental salaries; and 3) direct departmental supplies, which included the costs of both variable and fixed supplies. Using the direct method, departmental overhead is allocated indirectly, but most departmental salaries and supplies are assigned directly to patients.

The Traditional Method

The traditional method uses the ratio of departmental cost to charges multiplied by the patient's recorded charges. The 1978 total departmental expense for each cost center, before allocation of hospital overhead, was obtained from the 1978 Medicare Cost Report. This amount was verified against fiscal year-end operating room departmental expense statements and was adjusted if necessary. Total charges for that cost center were also obtained. These amounts were used to calculate the cost-to-charges ratio for each cost center. These ratios were multiplied by the charges recorded on each patient's bill to determine an adjusted cost estimate (Fig. 2). Cost estimates from each department were added to produce each patient's total costs. With the traditional method, all departmental costs are allocated indirectly, using measures such as operating room time or the average cost of diagnostic procedures.

Statistical Analysis

The two-tailed Student's t-test for paired samples was used to test for significant differences between the costs determined by the two methods when applied to the same patients. Significance was accepted if p was less than or equal to 0.05.

Results

Within each DRG, the total cost and the costs for each cost center determined by the direct method were significantly lower

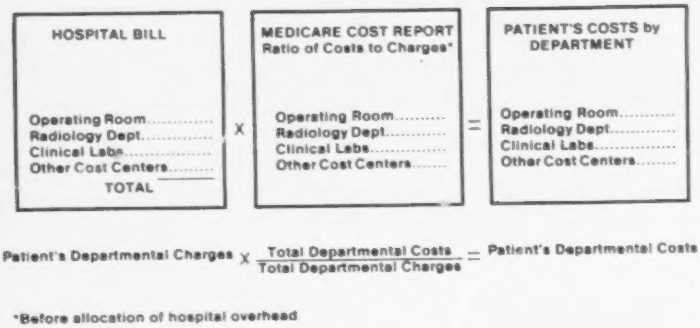

HOSPITAL BILL
Operating Room
Radiology Dept.
Clinical Labs
Other Cost Centers
TOTAL
X
MEDICARE COST REPORT
Ratio of Costs to Charges*
Operating Room
Radiology Dept.
Clinical Labs
Other Cost Centers
=
PATIENT'S COSTS by DEPARTMENT
Operating Room
Radiology Dept.
Clinical Labs
Other Cost Centers
Patient's Departmental Charges X Total Departmental Costs / Total Departmental Charges = Patient's Departmental Costs
*Before allocation of hospital overhead

FIG. 2. The traditional method for cost allocation.

than the costs determined by the traditional method (Tables 1 and 2). The costs for patients in DRG 198, who did not have major secondary diagnoses (Table 1), are seen to be lower in all three departments than those for patients in DRG 199, who did have major secondary diagnoses (Table 2) when the costs in one table are compared with the costs in the other table determined by the same method. The major determinant of differences in operating room cost, which contributed an important part of the differences in total costs between the two DRGs, was the duration of anesthesia time, which varied from 30 to 210 minutes.

Discussion

Two different cost-allocation methods were used to assign the costs from three departmental cost centers to patients who underwent inguinal hernia operations. The traditional method is the method that is used almost universally, with modifications, for allocating costs under traditional accounting systems and under the newer case-mix accounting systems. The direct method differs from the traditional method because it uses more direct measures to allocate the costs of services to patients who used those services.

The direct method that was used to assign operating room costs differed substantially from the traditional method. Using the direct method, many operating room services were associated directly with inguinal hernia patients, and the number and the costs of these services were determined separately from the number and the costs of services provided to all other patients. Using the traditional method, the costs of the services provided to inguinal hernia patients were first averaged with the costs of the services provided to all other patients and then allocated using a measure (operating room time) that only indirectly considered the number of services provided to inguinal hernia patients.

TABLE 1. Comparison of Mean Costs Determined by Two Methods of Assigning Costs to Patients in DRG 198 (Inguinal Hernia, Without Major Secondary Diagnosis)

Method	Total Cost	Operating Room Cost	Radiology Cost	Clinical Laboratory Cost
Direct	\$123.51 (36.88)*	\$ 89.76 (12.94)	\$7.43 (3.12)	\$26.32 (10.40)
Traditional	\$184.97 (17.49)	\$143.57 (32.58)	\$9.40 (7.04)	\$32.00 (13.37)

* Standard error in parentheses.

The direct method that was used to assign the costs of services from the clinical laboratories and the radiology department differed from the traditional method only in the way the cost of each service was determined, not in the way the number of services was counted. In both methods, simple examination of existing records was used to determine exactly which and how many diagnostic procedures were performed for each patient. By the direct method, the cost of each procedure was determined from existing measures of the amounts of time and supplies that were used to conduct the procedure. By the traditional method, the cost of each procedure was determined by less exact measures that were developed to insure accurate recovery of overall departmental costs and thus relied more on average costs for categories of procedures.

To determine the costs for each DRG, some costs are variable and direct, and the number used for each DRG can be counted easily from existing records, as was done here for diagnostic procedures. All that needs to be done to assign accurate costs for these services to each DRG is to measure each separate cost accurately. The traditional accounting method treats some costs that are variable and direct as if they were shared costs. To assign accurate costs for these services to each DRG requires first that the services be separated and accurately counted, and then that their separate costs be accurately measured, as was done here for many operating room services. However, some costs are truly fixed, indirect and shared, such as operating room overhead, and their allocation requires special methods that are described below.

This pilot study looked at only part of the costs for a limited number of patients in one hospital. The results should not be broadly generalized. Nevertheless, they illustrate that, at least in some situations, there are important differences in cost when the two different methods are used. In this pilot project, the direct method identified costs that were uniformly lower than the costs identified using the traditional method.

These results strongly suggest that some cross-subsidization was occurring with the traditional method. It is hypothesized that inguinal hernia patients, who receive relatively standard care, are charged more than their care costs to subsidize other patients who receive complicated care and are charged less than their care costs. The hospital might set charges to encourage cross-subsidization for a variety of reasons. The hospital might, for example, determine that some uncommon, complex procedures are so expensive that they could not be paid for by the patients who need them unless they were supported by relatively small subsidies drawn from the many patients who required less complex and expensive procedures. Alternatively, the hospital might set charges for emergency room services below cost, thus encouraging use of the emergency room earlier in the patient's ill-

TABLE 2. Comparison of Mean Costs Determined by Two Methods of Assigning Costs to Patients in DRG 199 (Inguinal Hernia, With Major Secondary Diagnosis)

Method	Total Cost	Operating Room Cost	Radiology Cost	Clinical Laboratory Cost
Direct	$201.74 (112.52)*	$118.06 (42.32)	$20.97 (26.68)	$62.71 (85.21)
Traditional	$310.43 (482.49)	$204.01 (98.27)	$30.94 (43.72)	$75.48 (102.77)

* Standard error in parentheses.

ness. The hospital would recover the difference from inflated charges to hospitalized patients, many of whom would have been admitted through the emergency room. In both cases, cross-subsidization may be in the public's interest and, therefore, be tolerated or even encouraged by third-party payors. Nevertheless, the information provided by the existing cost-allocation method is misleading.

It is possible, but less likely, that the total of all charges to inguinal hernia patients accurately reflects their total costs because the overcharges observed during this study are balanced by undercharges for services not examined. In this case, the information provided by the existing cost-allocation method is also misleading.

The evidence for cross-subsidization might have been stronger if the same methods had been used to measure costs for patients in a high-cost DRG that involved complex procedures or in some other DRG in which inflated charges might be expected. This was not done, however, because of limited time and resources.

These results also illustrate how the choice of a cost-allocation method could affect the operation of case-mix hospital accounting systems. Some cost-allocation methods obscure the relationship between patients and the resources they consume, and other methods describe the relationship more clearly. While case-mix accounting systems provide improved definition of hospital products in any case, it is only when the cost-allocation method accurately links resource consumption with those products that the full benefit of the system can be realized.

Even with the direct method, some costs, such as operating room overhead, are fixed, indirect and shared costs that cannot be allocated easily. Operating room overhead is a fixed cost; that is, a cost per patient cannot be computed without considering patient volume. It is also a shared cost that cannot be easily separated and assigned to the activities that cause it to arise. Since the cost is indirect, there is also the problem of associating the cost with particular patients in DRGs. If eight DRGs all make use of the operating room, it is difficult to decide how to allocate costs among the eight. A proponent for each DRG in turn could point to the other seven DRGs and note that the operating room would be needed for those seven, even if the eighth did not require it. Therefore, its use by the eighth DRG imposes no additional cost, and so no cost should be assigned to the eighth DRG for use of the operating room.

For decision-making purposes, the correct theoretical treatment of shared costs, such as those for the operating room, is to ignore them, since accounting theory calls for marginal cost pricing for optimal resource allocation. Thus, theoretically one should not be concerned with assigning shared costs to any particular DRG. The cost would be incurred even if the DRG were not offered. Obviously, however, this does not help the hospital administrator determine costs for pricing and reimbursement. All costs must be assigned to DRGs, or the hospital will not be able to recover all of its costs.

There are seven solutions to this allocation problem: 1) cause and effect; 2) ability to bear; 3) relative prices; 4) fairness for pricing; 5) equity; 6) physical units; and 7) benefits available.

With *cause and effect*, one assigns to a DRG the costs that it causes to be incurred. This method is preferred if a cause-and-effect relationship can be found, but, unfortunately, this is often not the case. When cause and effect cannot be determined, more arbitrary allocation bases are required.

Ability to bear is a frequently used basis for cost allocation in profit-oriented businesses. It assumes that some products can better bear the cost of large allocations. It allocates a large share of cost (e.g., the cost of the corporate president's salary) to the most profitable products or to those with the largest sales. The extra cost is a smaller proportionate burden for these products; the idea is similar in philosophy to that of a progressive income tax. Allocating costs to DRGs according to their ability to bear would continue the cross-subsidization that is a disagreeable feature of the current cost-allocation systems.

When prices are known, costs are often allocated based on the *relative prices* for each product. Thus, if the outside marketplace has decided that one DRG is worth twice as much as another, then one assigns two shares of joint cost to the first and one share to the second. Since hospital prices are not based on the consumer's utility but are based on costs measured in one way or another, assigning hospital costs based on relative prices is circular reasoning.

When outside market prices are not known, cost allocation is sometimes based on *fairness for pricing*. The goal of this method is to allocate costs in a way that ultimately achieves a fair set of prices. This method is commonly used for defense contracting, where the issue is to determine which costs will be borne by products sold to the government under a cost-plus contract and which by products sold to other buyers. It is, however, extremely difficult to determine what constitutes a fair set of relative hospital prices. In the absence of a generally acceptable standard of fairness, this method should not be used for hospital cost allocation.

The same problem exists when relying on the general concept of *equity*. Unfortunately, different people will have different concepts of what allocations are equitable. For example, if certain DRGs are related to patients' ages, one might expect Blue Cross and Medicare to differ in their opinion about an equitable allocation of costs to those DRGs because their clients have different characteristics, especially with respect to age.

Costs could be assigned on the basis of *physical units*, for example, the number of admissions for each DRG or the number of patient days. While the physical unit selected would clearly be arbitrary, at least it could be measured. This method may be useful as a last resort when other alternatives fail.

Benefits available is the final measure. For example, a stand-by defibrillator's cost could be assigned to patients who actually make use of the defibrillator in proportion to their use, as is done under cause and effect. Alternatively, one could argue that it was available for the use of patients in all DRGs or in some group of DRGs. An argument could be made for sharing costs among all patients who had the equipment available for their use. The authors propose that when direct cause-and-effect relationship does not exist, the preferred method for cost allocation is benefits available, although they did not use this method to allocate the shared cost of operating room overhead.

Since the direct cost-allocation method was not used in this study to measure all the costs of all the patients in each department, it was not possible to check for overlooked costs by comparing the sum of all patients' departmental costs with the department's overall costs. Therefore, the possibility exists that some costs were inadvertently excluded with the direct method, which would have lowered the costs measured with the direct method and could have explained the differences observed between the two methods. The authors consider this an unlikely explanation because differences between costs using the two methods were relatively uniform in all three departments and because of the care that was taken to avoid this problem when collecting information, especially in the operating room.

This pilot project did not provide information about the cost of collecting the additional information needed to allocate costs directly to patients under the new direct cost-allocation method. Whether the collection of additional information will substantially increase cost, and whether the increased cost will be offset by the benefits of a more accurate cost-accounting system, are questions that should be answered by larger and more comprehensive research projects. Ernst and Whinney, a major accounting firm, has developed a cost-allocation system under the sponsorship of the Illinois Hospital Association, similar to the one the authors propose,[22] and is presumably testing it for cost-effectiveness. Fetter has used a direct method to allocate costs under a case-mix cost-accounting system in the Department of Radiology at the Yale-New Haven Hospital.[18] She was able to determine which year-to-year variances in cost were the result of changes in the department's operating efficiency, which the result of changes in patient mix, and which the result of changes in the clinical practices of the medical staff. She concludes that such a powerful accounting system ought to be worth its cost but does not provide information about the actual cost of the system.

Finally, improved cost allocation could lead to improvement in the case-mix measures themselves. Diagnosis-related groups, for example, are formed by identifying patients with similar diagnoses and lengths of hospital stay, using length of stay as an easily determined but indirect measure of the bundle of services used by patients in the DRG. Several investigators are attempting to develop improved classification systems in which patients with similar diagnoses have similar costs instead of similar lengths of stay, because cost should be a more accurate measure of resource consumption.[23] These efforts, however, depend on existing cost-allocation methods, with their problems of inaccurate information and cross-subsidization. Further improvement would be expected if case-mix measures could be developed using data from a more accurate cost-accounting system with an improved cost-allocation method such as the one proposed in this article.

Conclusions

The current DRG cost-accounting system should be refined to provide a clearer assessment of 1) which costs should be assigned directly to patients in specific DRGs; 2) which costs are fixed, indirect and shared costs that require allocation; and 3) which allocation method and bases should be used.

Clearly, there is no one method for allocating costs to DRGs. Costs that are variable, direct and separable should be assigned to specific DRGs using simple measures of the number of services already available in the medical record and more accurate measures of the cost of each service.

At present, there are alternative ways of assigning other costs. Accuracy will be improved if more costs that are treated as fixed and indirect costs in traditional accounting systems are assigned instead to specific DRGs using methods based on cause and effect.

We propose allocating truly joint costs using the method of benefits available. Alternatively, it may be possible to choose one of the other methods we describe, at least in special situations, if there is a defensible rationale for choosing that method over the others.

The use of more accurate methods of cost allocation will serve several purposes. It will assist hospital administrators in determining whether charges accurately reflect the cost of the resources used to care for individual patients. It may be used by third-party payors to establish equitable reimbursement schemes that do not penalize hospitals that care for more complex patients who require more costly services. It may provide better information for regional planning. It should provide hospital managers with one of the instruments they need to improve planning, budgeting and control of hospital operations. It may lead to improvement in the case-mix measures themselves.

Acknowledgment

Mark S. Levitan, Executive Director of the Hospital of the University of Pennsylvania, helped identify the issues discussed in this article, served as supportive consultant and adviser during the research, and provided valuable perspective while the article was being written.

References

1. Holden WW. Hospital budgeting: state of the art. Hospital and Health Services Administration 1978;23:51.

2. Drebin ME. Financial information systems: the key to hospitals' survival. Hospitals 1978;52:88.

3. Griffith JR. Measuring Hospital Performance. Chicago, Ill.: Inquiry, 1978.

4. Finkler SA. Cost-effectiveness of regionalization: the heart surgery example. Inquiry 1979;16:264.

5. Thompson JD, Fetter RB, Mross CD. Case mix and resource use. Inquiry 1975;12:300.

6. Mills R, Fetter RB, Riedel DC, Averill RR. AUTOGRP: An interactive computer system for the analysis of health care data. Med Care 1976;14:603.

7. Fetter RB, Thompson JD, Mills RE. A system for cost and reimbursement control in hospitals. Yale J Biol Med 1976;49:123.

8. Fetter RB, Mills R, Riedel DC, Thompson JD. The application of diagnostic specific cost profiles to cost and reimbursement control in hospitals. J Med Syst 1977;1:137.

9. Thompson JD, Averill RF, Fetter RB. Planning, budgeting and controlling: one look at the future. Case mix cost accounting. Health Serv Res 1979;14:111.

10. Fetter RB, Shin Y, Freeman JL, et al. Case mix definition by diagnosis-related groups. Med Care 1980;18(June supplement):1.

11. Bentley JD, Butler PW. Describing and Paying Hospitals: Developments in Patient Case Mix. Washington, D.C.: Association of American Medical Colleges, 1980.

12. Young WW, Swinkola RB, Hutton MA. Assessment of the AUTOGRP patient classification system. Med Care 1980;18:228.

13. Horn SD, Schumacher DN. An analysis of case mix complexity using information theory and diagnostic related grouping. Med Care 1979;17:382.

14. Luke RD. Dimensions in hospital case mix measurement. Inquiry 1979;16:38.

15. Reiss JB. A conceptual model of the case-based payment scheme for New Jersey hospitals. Health Serv Res 1980;15:161.

16. A Prospective Reimbursement System Based on Patient Case-Mix for New Jersey Hospitals, Third Annual Report, Volume I. Trenton, New Jersey: New Jersey State Department of Health, 1980. (Mimeographed.)

17. Grimaldi PL. Equity and efficiency implications of case-mix reimbursement in New Jersey. In: Glandon GL, Shapiro RJ, eds. Profile of Medical Practice, 1980. Monroe, Wis.: Center for Health Services Research and Development, American Medical Association, 1980; 81-90.

18. Fetter AL. Cost Measurement and Control in a Hospital Patient Service Department: A New Approach. Thesis, Department of Epidemiology and Public Health, Yale University, 1980; 131.

19. Murphy CM. Comparison of Cost-Finding Methods and Analysis of Variables Affecting Resource Consumption with a Diagnostic Related Groups. Thesis, University of Minnesota Program in Hospital and Health Care Administration, 1980; 79.

20. Laboratory Workload Recording Method, 1980 Edition. Traverse City, Michigan; College of American Pathologists, 1980.

21. Horngren CT. Cost accounting: a managerial emphasis. 4th edition. Englewood Cliffs: Prentice-Hall, Inc, 1977.

22. A Project to Design and Pilot Test a Diagnosis-Related-Patient Care Unit System, Executive Summary. Chicago; Ernst and Whinney, 1980.

23. Schumacher D, Horn S. Isocost groups. In: Bentley JD, Butler PW. Describing and Paying Hospitals: Developments in Patient Case Mix. Washington, D.C.: Association of American Colleges, 1980; 10–11.

MEDICAL CARE
May 1982, Vol. XX, No. 5

Three Case-Type Classifications: Suitability for Use in Reimbursing Hospitals

RICHARD P. AMENT, A.M., JANICE L. DREACHSLIN, PH.D., EDWARD J. KOBRINSKI, M.A., AND WALTER R. WOOD, A.M.

This study compared three case-type classifications—the cross-classification of the Commission on Professional and Hospital Activities, Diagnosis-Related Groups (DRGs) and Staging—with respect to per cent of variance in total patient charges accounted for. The purpose was to assess the relative usefulness of the classifications for application in hospital reimbursement schemes. The sample consisted of 50 hospitals. A nested analysis of variance was performed with case type nested within hospital. Per cent of variance accounted for was calculated for each of three data sets: the full data set, a truncated version of that set and a logarithmically transformed version. Results support the contention that none of the currently available classifications accounts for enough variance to permit straightforward use of case-type standard costs in a reimbursement mechanism. New developments in case-type classification may result in a classification that is more suitable for this use.

INFLATION-ADJUSTED EXPENDITURES for hospital goods and services have risen at an average annual rate of 8 per cent since 1960.[1] During that same period, proportionately more health care has been paid for by third-party payors, including federal and state governments through Medicare and Medicaid programs. Concerned third-party payors have begun to experiment with alternative reimbursement schemes in attempts to slow the steady increase in expenditures for hospital goods and services.

With the hope of eliminating inflationary incentives in retrospective reimbursement methods, experiments with prospective reimbursement schemes are being conducted in which the reimbursement rates for hospital goods and services delivered to patients are set prior to the period in which costs are incurred. Research results indicate that selected prospective schemes are experiencing some success in containing health care expenditures.[2] The National Hospital Rate Setting Study, now in progress, will attempt to identify characteristics of effective prospective reimbursement mechanisms.

Any reimbursement scheme should set standards for the allowable cost of goods and services, preferably taking case mix into account. Case mix is defined as the proportion of patients falling into mutually exclusive case types. The per-case system of taking case mix into account is to set a separate standard for each case type in some classification of patients and to reimburse accordingly. The equity of this

---

From the Research and Statistics Department of the Commission on Professional and Hospital Activities, Ann Arbor, Michigan.

This research was funded by the Health Care Financing Administration, Department of HHS, under Contract 500-78-0002. The views expressed are those of the authors.

Requests for reprints: Richard P. Ament, Commission on Professional and Hospital Activities, P.O. Box 1809, Ann Arbor, MI 48106.

0025-7079/82/0500/0460/$00.90 © J. B. Lippincott Co.

method depends in part on the characteristics of the case-type classification used.

We have argued elsewhere that a case-type classification for use in hospital reimbursement should have a number of specific qualities.[3] The quality of major importance in this discussion is economic meaningfulness. By this is meant homogeneity of patients' resource needs within a case type.

Using the charges for resources delivered to a patient as a proxy for the patient's resource needs, this study empirically evaluates the economic meaningfulness of three available case-type classifications: the cross-classification of the Commission on Professional and Hospital Activities, Diagnosis-Related Groups (DRGs) and Staging. More equitable reimbursement patterns will result when a case-type classification that exhibits a relatively high degree of economic meaningfulness is employed. None of the studied case-type classifications was developed with economic meaningfulness as the sole characteristic of importance, and all of them aggregate patients into broad diagnosis groups that are subsequently stratified.

The cross-classification of the Commission on Professional and Hospital Activities, in its original version, aggregates *H-ICDA-2* principal diagnosis codes into the 349 categories of CPHA List A: Hospital Diagnosis Groups, each of which is further categorized to define case types. Stratification into case types is symmetric across diagnosis groups. In this study, each diagnosis group is divided into ten categories based on presence or absence of operations, and five broad age groups. Stratifiers were selected because of their relationship with length of stay and severity of illness. List A was developed for use in the review and evaluation of care of hospitalized patients; the further stratification, supplemented by a split on presence or absence of secondary diagnoses, has been used for presentation and analysis of length-of-stay statistics.[4]

Diagnosis-Related Groups aggregate patients in such a way that any case type is reasonably homogeneous with respect to length of stay. An interactive statistical analysis system, AUTOGRP, was used to partition patient data, using length of hospital stay as the dependent variable and a collection of independent variables similar to those of the Uniform Hospital Discharge Data Set. Variables selected to stratify Diagnosis-Related Groups are as follows: primary diagnosis, secondary diagnoses, primary surgical procedure, secondary surgical procedures, age, and the presence or absence of a psychiatric service. All diagnosis groups were initially aggregated by clinicians into 83 broad, mutually exclusive, major diagnostic categories. Further division of each major diagnostic category into specific case types was an iterative process, and decisions were based on a number of factors, including the reduction in the within-group sum of squares attained by the grouping, the number of groups and clinical decisions. Diagnosis-Related Groups were developed as an aid in utilization review.[5]

Staging formed original diagnosis groups, called disease conditions, by identifying the organ system affected by the disease and the cause of the disease. A panel of physicians defined the disease conditions, each of which was subsequently stratified into case types on the basis of severity of the illness. The staging method subdivides each of the disease conditions into three basic stages of disease severity:

Stage I:	Condition with no complications, or problems of minimal severity.
Stage II:	Condition with local complications, or problems of moderate severity.
Stage III:	Condition with systemic complications, or problems of maximal severity.[6]

The present study uses patient total charges as a proxy for their ideal total costs. That is, a given case-type classification is considered to exhibit economic meaningfulness in the degree to which the classification accounts for the variance in charges as opposed to costs. The appropriateness of this substitution is discussed below.

Not addressed in this study is the usefulness or equity of specific categories within each classification. Also not addressed are the issues of how hospitals should be grouped for the purpose of determining and applying the standard costs, how the standard costs should be compiled and how they should be used in calculating reimbursements.

Methods

To assess the technical appropriateness of a given classification for use in a prospective reimbursement scheme, one would ideally go through the following process:

1. Obtain patient records on a large, representative sample of admissions to hospitals of a given class, that is, that have similar staff resources, equipment availability and unit costs.
2. For each patient, determine the treatment of choice, considering both costs and benefits.
3. For each patient, calculate the cost of such treatment in an ideal institution of the given class.
4. Measure the homogeneity of these ideal costs within case type.

But since information on the amount of resources appropriate for the care of individual patients is generally not available, one might take, as a proxy, the amount of resources historically used for the care of individual patients. And since resource consumption figures themselves are not generally available for individual patients, charge figures were used in this study as second-order proxies for the corresponding appropriate levels of resource consumption.

Three kinds of errors are introduced by this substitution:

1. The bundle of resources actually used in the care of a given patient may be different from the ideal bundle. In fact, some patients would ideally have been treated in ambulatory settings rather than as inpatients.
2. The charges for the various goods and services supplied to a given patient may not accurately mirror the corresponding actual costs.
3. The actual costs, in turn, may not coincide with the costs for the same goods and services in an ideal hospital of the given class.

It may be seen that interhospital variation in patient management, charge structure and unit costs could, through its effect on the three kinds of errors, substantially inflate the measured heterogeneity within case type. In this study, the interhospital source of variation is eliminated by taking the measurements within hospitals. But the available data did not afford a method of reducing the remaining error variation, a large part of which may be due to within-hospital differences in the management of patients of the same kind. The analysis in this study assumes that the remaining error variation inflates the heterogeneity within case type equally among the three classifications. There appears to be no reason to suspect that this assumption is seriously in error.

Data Files

The data that were analyzed came from hospitals participating in the Study of Patient Charges of the Commission on Professional and Hospital Activities (CPHA). The Study of Patient Charges is an optional extension of the Professional Activity Study (PAS) in which the hospital adds each patient's total charge to the standard PAS data set and receives special reports analyzing these charges. Study of Patient Charges hospitals are a self-selected group of nonfederal, short-term hospitals in the

United States characterized by a wide variation in size, teaching status and region of the country.[3] Hospitals included in this study were the 50 that participated in the Study of Patient Charges during their entire 1976 fiscal year and that signed releases of their patient data for that time period for use in the study. The resulting file of base data contained 494,523 discharges. For each hospital in this group, all patients discharged during the 1976 fiscal year were included in the study with the following exceptions: a) births, excluded because charges recorded on births are unreliable since charges for newborn babies are often billed to the mothers; b) patients with invalid or missing data for age, principal diagnosis, length of stay or total charge (a charge was considered invalid if the average charge per day fell outside of wide, preset ranges).

Three additional data files were created from the base data file—a staged file, a trimmed file, and a file that was both staged and trimmed. The two staged files included only disease conditions for which a staging algorithm was then available; approximately 46 per cent of the patients in the base file were staged. The two trimmed files excluded patients who died in the hospital, were transferred to another hospital or stayed for more than 70 days.

The untrimmed files were used for analysis because reimbursement schemes would avoid a certain arbitrariness and complexity in application by minimizing exceptions, and it would be useful to know how the classifications would perform in such schemes. The trimmed files were also used because some current schemes do in fact trim the data when compiling weights and when using the weights in determining reimbursement amounts.[7]

Statistical Model

The task was to measure, for each case-type classification, the homogeneity of patient total charges within case type. Homogeneity, though, is traditionally expressed in terms of its opposite, heterogeneity, as manifested by the variance. To estimate this variance free from the confounding hospital effect, and to provide an estimate of total variance within hospital for comparison, a nested analysis of variance was employed for each classification. A full two-way analysis of variance was not done because there were too many levels in each of the explanatory variables, hospital and case type, to make it feasible. The model used for the nested analysis was:

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk},$$

where Y_{ijk} = total charge for patient k in case type j in hospital i; μ = overall average of charge; α_i = average charge for hospital i, minus μ; $\beta_{j(i)}$ = average charge for case type j within hospital i, minus $(\mu + \alpha_i)$; and ϵ_{ijk} = charge of patient k in case type j in hospital i, minus $(\mu + \alpha_i + \beta_{j(i)})$. This model was fitted by least squares, yielding a mean square for hospitals; the mean square for case types within hospital, MS_T; and the mean square for patients within case type within hospital, MS_E.

The nature of the nested analysis may be clarified by the following discussion. If there were only a single hospital in the data set, one would obtain the case-type mean square MS_T and the residual mean square MS_E from the fitted model, and then calculate the corresponding variance components by the formulae:

$$\sigma_T^2 = (MS_T - MS_E)/n_0,$$
$$\sigma^2 = MS_E,$$

where n_0 = the weighted mean number of patients per case type

$$= (\Sigma n_j - \Sigma n_j^2/\Sigma n_i)/(t - 1),$$

where n_j = the number of patients in case type j and t = the number of case types represented by at least one patient.[8]

Since more than one hospital was included in the data, the estimates σ_T^2 and σ^2 were calculated as the appropriately weighted mean of the individual hospital estimates. By thus nesting case type within hospital, the analysis separated the confounding hospital effect. The NESTED procedure of the Statistical Analysis System (SAS) was used to perform these calculations.

Logarithmic Transformation

The analyses were performed using the charges just as they were found in the input data. Additionally, analyses were performed on a transformation of those charges in the two untrimmed data files. A logarithmic transformation was used to reduce the extremely large sampling error inherent in the highly skewed distributions characteristic of patient costs and charges, and to make additive the hospital effect, which may be multiplicative in the untransformed data. A pure logarithmic transformation, however, exaggerates variation among charges at the lower end of the distribution and somewhat distorts analyses purporting to compare variations in the charges themselves. Both disadvantages were reduced by adding an arbitrary constant, $200, to each charge before taking the logarithm. The $200 figure was suggested by some experience with length-of-stay data, where adding a day or two to the stay tended to make the logarithms normally distributed; $200 is a rough dollar equivalent of one or two hospital days in 1976 data.

An additional benefit of performing the parallel analyses was the procurement of evidence for determining whether small changes in the distribution of costs would affect the relative economic meaningfulness of the three classifications.

Results

Table 1 gives the results from the nested analyses of variance of patient charges. For ease of interpretation, the statistic tabulated for each analysis is the proportion of variance accounted for by case type within hospital. This statistic may be interpreted as a measure of relative homogeneity within case type. Little difference was found between the CPHA cross-classification and the DRG classification in this measure.

The CPHA cross-classification and the DRG classification each accounted for approximately a third of the within-hospital variance of patient charges in the full base data, which included all patients except newborn and a small contingent of patients who were excluded for missing data elements or unreasonable charge figures. These two classifications each accounted for nearly 40 per cent of the within-hospital variance when the data set was trimmed, i.e., when deaths, transfers and patients staying for more than 70 days were excluded. In the untrimmed data set, when the logarithmic transformation was used to reduce the erratic effect of occasional extremely high charges, the two classifications accounted for about half of the within-hospital variance.

When the data set was reduced to include only patients who belonged to a staged disease condition, the DRGs and the CPHA cross-classification were substantially more successful than the staging classification at accounting for variation in total charges. But when the case types of the staging classification were each split into two subtypes, for operated and nonoperated patients, the resulting classification was almost as successful as the DRGs and the CPHA cross-classification.

Discussion

The fact that the CPHA cross-classification and the DRGs had approxi-

TABLE 1. Per Cent of Within-Hospital Variance of Individual Patient Total Charges Due to Case Type*

	Full Base Data		
	Untrimmed Data (N = 494,523)		Trimmed Data† (N = 476,926)
Classification	Variable Studied = Charge	Variable Studied = log (200 + Charge)	Variable Studied = Charge
CPHA cross-classification	33.0%	50.3%	38.8%
DRGs	31.5	47.2	39.9

	Staged Patients Only		
	Untrimmed Data (N = 201,767)		Trimmed Data† (N = 195,315)
Classification	Variable Studied = Charge	Variable Studied = log (200 + Charge)	Variable Studied = Charge
CPHA cross-classification	40.4%	56.4%	45.6%
DRGs	40.5	55.0	48.9
Staging	29.5	44.1	34.8
Staging × operated/ nonoperated	38.9	51.9	43.7

* The quantity tabulated is $\sigma_T^2/(\sigma^2 + \sigma_T^2)$, where σ_T^2 = estimated variance component due to case type within hospital and σ^2 = estimated variance component due to error, i.e., residual variance.

† The trimmed data excludes deaths, patients transferred to other hospitals and patients staying for more than 70 days.

mately the same overall heterogeneity within case type suggests that the use of either classification would result in approximately the same payments to hospitals in a prospective reimbursement scheme. This expectation is corroborated by a finding in another study in which a correlation of r = 0.97 was obtained between case-mix indices based on the two classifications. One index was constructed from the CPHA cross-classification, using charge weights compiled from Study of Patient Charges data, the other from DRGs, using New Jersey standard cost weights. The study data represented one year's discharges from 316 short-term hospitals in the United States; each index value was calculated for the totality of discharges from one hospital.[9]

The greater heterogeneity within case type in the staging classification suggests that reimbursements based on that classification would be less equitable than reimbursements based on either of the other two classifications, although the disadvantage would be slight if the operated/nonoperated split were added.

It may be noted that σ_T^2 is the combined result of the case-type main effect and the interaction between case type and hospital. Part of the interaction should rightfully be allocated, in this study, to the within–case-type heterogeneity, since the set of case-type weights to be used in a reimbursement formula would be the same for all hospitals in any hospital class. But since hospitals have not been classified for the purpose of developing sets of standard costs to use in the reimbursement mechanism, the proportion to be allocated could not be determined, and, in any case, an estimate of the interaction is not available from the nested model. Inability to make this allocation, however, affects all classifications in the study and is thought to add little bias to the comparisons among them.

As noted above, the observed variances of charge within case type, large by any standard, would be still larger if they included interhospital differences, even if pricing differentials were not a factor. These large variances indicate that a standard cost allowance for a given case type in a particular hospital will often differ markedly from the ideal cost of the individual patients to whom it applies. As a result, a reimbursement scheme using any of these classifications would be inequitable if applied in a straightforward manner. The current New Jersey prospective reimbursement project greatly reduces this inequity by paying according to the weighted average of the state standard and a figure derived from the hospital's previous actual costs, giving the state standard a low weight in any case type in which the variance of costs in base data is high.[7]

The high variances of charges within case type appear to be due chiefly to the following causes. Many case types cover a wide range of medical conditions, whose treatments vary greatly in resource requirements. Patients with a given condition may vary in their reason for admission, e.g., for diagnosis, treatment or terminal care, and the cost of their management will accordingly vary greatly. A few patients with unusually severe manifestations of their disease, with unusual complications or with other factors causing extra expense, can add substantially to the variance. That the latter cause is an important influence is demonstrated by the drop in variance observed in Table 1 when the full patient file is replaced by the trimmed file, from which patients staying more than 70 days were excluded, or when the effect of extremely large charges is reduced by the use of the logarithmic transformation. Additional causes of high heterogeneity would be management and price differentials, and inaccuracy or incompleteness in the medical record, or in the coding and abstracting of the medical record data.

What then are the prospects of achieving a patient classification more satisfactory for use in hospital reimbursement? Developments in this direction, occurring since the present study was done, include the following:

1. DRGs have been redefined in terms of the *ICD-9-CM* diagnosis and operation codes. A larger and more nationally representative patient sample was used as the data base for the AUTOGRP process, to which a larger number of physicians gave input. But the limitation of the number of terminal groups to a few hundred will make serious heterogeneity in a number of groups unavoidable.

2. The CPHA cross-classification has been redefined in terms of *ICD-9-CM*, using 14 per cent more diagnostic categories, but the basic structure remains unchanged and likely would show little reduction in heterogeneity of patient costs or charges within case type.

3. The staging classification has been extended to cover all remaining disease conditions. But the overall heterogeneity within case type in the completed classification is not expected to be much, if any, smaller than that found here. This is because nearly half of short-term general hospital patients are already accounted for in the studied case types and because the remaining case types include residual categories. Overall, however, if the case types formed by the different stages of the disease conditions are further subdivided according to purpose of admission and management strategy, the heterogeneity within case type is likely to fall much lower than any variance found in this study. Support for this conjecture is provided by the substantial drop in variance achieved by simply splitting the case types into cells for operated and nonoperated patients.

4. The generalized management paths being developed under the direction of Young at Blue Cross of Western Pennsylvania also promise to result in much lower variances than those found in this study because of the detailed attention paid to clinical attributes, with emphasis on reason for admission to which costs are sensitive.[10]

5. A four-level illness-severity classification developed by Horn and her colleagues at the Johns Hopkins University has been tested on each of six diseases in three hospitals, and has yielded lower variances of length of stay and of patient total charge than DRGs, staging or generalized

management paths, according to a personal communication from Dr. Horn. Assuming that the further testing now under way should bear out the early promise of this classification, a possible impediment to its general acceptance for use in hospital reimbursement is that its application requires some use of judgment.

Conclusion

Diagnosis-Related Groups and the cross-classification of the Commission on Professional and Hospital Activities showed little difference in within–case-type heterogeneity of patient charges used as proxies for costs. The high levels of heterogeneity within case type preclude full use of cost standards for the case types of these classifications in hospital reimbursement schemes. Potential for the achievement of a classification that would exhibit less within–case-type heterogeneity of charges or costs is promised by current methodologic developments in the identification of disease stages, generalized management paths and illness-severity levels. Such a classification would make possible more equitable payments to hospitals under reimbursement schemes that take case mix into account.

References

1. Health Care Financing Administration. First annual report of the National Hospital Rate-Setting Study: A Comparative Review of Nine Prospective Ratesetting Programs. Baltimore: Health Care Financing Administration, Office of Research, Demonstrations, and Statistics, August 1980.

2. Biles B, Schramon CJ, Atkinson JG. Hospital cost inflation under state rate-setting programs. N Engl J Med 1980;303:664.

3. Wood WR, Ament RP, Kobrinski E. A foundation for hospital case mix measurement. Inquiry 198; 18:247.

4. Commission on Professional and Hospital Activities. Length of Stay in PAS Hospitals, by Diagnosis, United States, 1976. Ann Arbor: Commission on Professional and Hospital Activities, 1977.

5. Fetter RB, Shin Y, Freeman JL, et al. Case mix definition by diagnosis-related groups. Med Care 1980;18(Feb suppl).

6. Garg ML, Louis DZ, Bliebe WA, et al. Evaluating patient costs: the staging mechanism. Med Care 1978;16:191.

7. New Jersey State Department of Health. Third annual report (Vol. I): A Prospective Reimbursement System Based on Patient Case Mix for New Jersey Hospitals, 1976–1983. Trenton: New Jersey State Department of Health, 1980.

8. Steel RGD, Torrie JH. Principles and Procedures of Statistics. New York: McGraw-Hill, 1960.

9. Watts CA, Klastorin TD. The impact of case mix on hospital cost: a comparative analysis. Inquiry 1980;17:357.

10. Young WW. Measuring the Cost of Care Using Generalized Management Paths, Year One. Pittsburgh: Blue Cross of Western Pennsylvania, 1979.

MEDICAL CARE
May 1982, Vol. XX, No. 5

Comparing Classification Methods: Measurement of Variations in Charges, Length of Stay, and Mortality

SUSAN D. HORN, PH.D.,* AND DALE N. SCHUMACHER, M.D., M.ED.†

We define and examine three alternative systems for categorizing hospital patients. The first system is based on discharge abstract data alone; the second system is based on discharge abstract data that has been reabstracted from charts for completeness and accuracy; the third system is based on a severity-of-illness index within a diagnostic group. Using acute myocardial infarction patients as an illustrative example, we examine the homogeneity of these categorization systems with respect to charges, length of stay and mortality rates. Our results indicate that a measure of patient severity of illness is essential to produce homogeneous categories to study hospital productivity for utilization review assessment by PSROs, for planning purposes by Health Systems Agencies, for hospital rate setting by cost review commissions and for internal hospital financial management and utilization review studies.

MANY AGENCIES now collect hospital discharge abstract data to be used by Professional Standards Review Organizations (PSROs), Health Systems Agencies (HSAs) and state hospital rate-setting regulatory bodies. But hundreds of thousands of patient discharge abstracts are not useful as an unedited mass of data; they must be summarized or categorized if they are to be useful in quality-of-care and utilization review assessment by PSROs, in planning by HSAs, and in hospital rate setting by cost review commissions.

What sort of case-mix summarization or categorization system is best for these data? This article examines in detail three alternative systems based on: 1) discharge abstract data alone (DRG); 2) discharge abstract data that has been reabstracted from charts for completeness and accuracy (MD-DATO); and 3) a severity indexing of the diagnostic group (AS-SCORE). We have chosen acute myocardial infarction as the illustrative diagnosis, and we compare cases from a university teaching hospital with cases from a nonteaching community hospital.

One intuitive characteristic that a good categorization system should have is *medical meaningfulness* of the groups: patients with similar medical problems should be grouped together. A second characteristic of a good categorization system is *homogeneity* within the groups with respect to such chosen attributes as length of stay, severity of illness, clinical characteristics, charges, etc. Two common measures of homogeneity of data are 1) small values of the *coefficient of variation* (ratio of the

---

* Department of Health Services Administration Johns Hopkins University, Baltimore, MD.

† President, Rockburn Institute.

Partially supported by Grant No. 18-P-97045 and 18-P-97117 from HCFA, Dept. HHS.

Reprint requests: Dr. Susan D. Horn, 624 N. Broadway, Room 326, Baltimore, MD 21205.

0025-7079/82/0500/0489/$01.10 © J. B. Lippincott Co.

standard deviation divided by the mean) and 2) large *reduction in variance*, i.e., the sum of squares in the subgroups are reduced as compared with the sum of squares in the initial group. A third characteristic of a good categorization system is that the inclusion into or exclusion from a group should depend on *intrinsic patient characteristics* and should not be influenced easily by physicians or hospital administrators. Variables such as "procedure or not," or "intensive care unit or not" can be influenced easily by physicians and hence should be used cautiously in forming a categorization system. Variables such as "age" and "interacting secondary diagnoses" are less subject to this sort of influence.

There exist many different variables with which to form a patient categorization system, but the best choice of variables depends on the ultimate use of the system. If the system is being developed for quality assurance purposes, one may wish to develop a categorization system in which the risk of inpatient death in each group is the same, assuming that the patients are receiving appropriate medical care.[1] If the system is being developed for examining length of stay (LOS), then one should have a categorization system such that the variation in LOS of cases in each group is minimized.[2] Or, if the system is being developed for reimbursement purposes, then one may wish to develop a categorization system in which the groups are determined so that the variation in charges or costs in each group is minimized.[3]

There are three categorization systems in use today. One is based on *ICDA codes* and contains as many groups as there are three or four digit ICDA codes (depending on which coding system is used). But this system produces about 1,000 groups (9,999 when the fourth digit code is used), which are very difficult to analyze in a comprehensive, coordinated fashion. A second system, the Professional Activity Study (PAS) A List, was formed by collapsing all the ICDA codes into 349 categories.[4] Each of these diagnostic categories is further divided into 20 subgroups, taking into account operation (yes or no), secondary diagnosis (yes or no), and age (five age groups). The PAS-A List has, therefore, almost 7,000 separate category subgroups.

A third common categorization system is the diagnostic-related grouping system (DRG) developed at Yale–New Haven Hospital.[2] This system has 383 groups that were formed from an original set of 83 major diagnostic categories (MDCs), such as acute myocardial infarction, congestive heart failure (CHF), pneumonia, etc. Each major diagnostic category was split into smaller subgroups using such variables as presence or absence of secondary diagnoses, presence or absence of procedures, and age, with the goal of producing medically meaningful groups with a minimum variation in LOS within each group. Evaluations of the DRG system can be found elsewhere.[5,6]

In this article we describe and compare several categorization systems. We focus on the disease and disease category *acute myocardial infarction* (AMI). For this condition, the ICDA code 410, the PAS A list grouping, and the DRG 121 contain exactly the same cases. In the DRG system, AMI is not only the major diagnostic category (MDC), but it is also the final terminal group (DRG). We shall use the DRG system as the standard, since it is widely used. We shall also describe two new categorization concepts, the physician-data-optimal (MD-DATO) system and a patient severity system.[7]

It is our intention to: 1) illustrate the strengths and weaknesses of discharge abstract data for cost reimbursement purposes; 2) compare the three systems—the DRG system, the MD-DATO system and the patient severity system—with respect to homogeneity of charges and LOS within each set of groupings and with respect to mortality rate analyses; and 3) distinguish between case mix and case-mix severity. In

BEST AVAILABLE COPY

order to make these comparisons, we use: 1) discharge abstract data aggregated in the usual way, as employed by regulatory agencies; 2) several refinements of this aggregation system; and 3) detailed chart data. We give a detailed comparison using data from a 300-bed community hospital with no teaching programs, as well as data from a large metropolitan 1,000-bed medical center teaching hospital with over 20 residency training programs. The disease AMI was selected because it is one of the most difficult diseases to separate into homogeneous subgroups using discharge abstract data. That is the reason there was only one DRG for AMI patients. Also, it is one of the most CHARGE- and LOS-variable diseases in the HCFA Medicare data base.

Methods

In this section we describe briefly the MD-DATO system and the patient severity index, using AMI as an example. As part of a project to refine the existing DRGs with charge per case as the dependent variable and with systematic physician input, a physician specialty panel of five cardiologists and one internist considered all six cardiology major diagnostic categories (hypertensive heart disease, AMI, ischemic heart disease, arrythmia and slowed conduction, heart failure, and carditis and valvular disease) used in the DRG system.

The physician panel recommended that the medical meaningfulness and charge homogeneity of the AMI category could be improved by taking into account those secondary diagnoses that were most likely to explain differences in charges and mortality rates among AMI patients. Subsequent statistical analyses, using the 1977 Maryland discharge abstract data, were performed on the physician-generated groups as well as on minor modifications (partial collapsing) of these physician groups to make them more homogeneous with respect to charge. It is these modified groups that are the MD-DATO groups in this study. The data required for the DRG analyses and the MD-DATO analyses come from patient discharge abstracts; hence, patients can be placed into respective categories via computer. However, in the present study, because of its evaluative nature, we allocated the patients to the MD-DATO groups by a thorough review of the chart to make certain that discharge abstract coding omissions would not adversely affect the analyses.

In the MD-DATO system, the AMI patients are divided into four subgroups. These subgroups, based on the secondary diagnoses of the patient, are: 1) shock; b) pulmonary embolism and infarction, systemic embolism, heart failure, third degree heart block, cardiac arrest; c) hypertension, other arryhthmia (atrial fibrillation, atrial flutter, paroxysmal atrial tachycardia, ventricular fibrillation, flutter and tachycardia), venous thrombosis, recurrent pain; and d) other secondary diagnoses. A full list of all the diagnoses is available from the authors upon request.

The index of inpatient severity of illness[7,9] was developed in response to the need to compare case-mix severity among physicians whose practice patterns led to a wide variation in their patients' LOS, charges and mortality.[10] The patient-severity system is a simple, rudimentary classification system for severity of illness of medical patients that is based on clinical data from patients' medical records at the time of discharge. Statistical analyses of this index are presented elsewhere.[7] Efforts are now underway to study a new severity measure that is also applicable to surgical, obstetric, gynecologic and pediatric patients.

The severity index is called "AS-SCORE," an acronym that was derived from the principal attributes of this index:

A–Age of the patient.

S–Systems involved (cardiovascular, pulmonary, etc.).

S–Stage of disease [acute, chronic, chronic progressive, chronic with exacerbation, advanced (nonprogressive, progressive)].

CO–COmplications.

RE–REsponse to therapy.

The attributes assessed are *patient characteristics* such as age; *illness characteristics* such as organ systems involved, stage of the disease and assessment of complications; *results of therapeutic interventions;* and *outcomes.* Because the elements considered in this severity index are independent of a specific disease, the AS-SCORE index can be used with many medical diseases, medical major diagnostic categories or medical DRGs.

After review of the patient's medical record at the time of discharge, the patient is placed into one of four severity classes. Classes 1 through 4 are defined as follows (key decision items vis-à-vis AMI are italicized):

Class 1 includes patients with single system involvement in which either the disease stage is acute and nonprogressive or there are multiple, minor chronic disease states that do not interfere with normal activities of daily living. The patients have *no complications,* and their *response to therapy* is *complete,* prompt and sustained. There are no class 1 AMIs, since residual disease remains and is not eradicated by the therapy.

Class 2 includes patients with single or multiple system involvement. The disease stage in this class is acute or chronic (a duration of six months or more) and/or progressive. The patient may have complications that are treatable *without severe residual impairment.* The response to therapy is either complete or controlled, prompt or delayed, or sustained or intermittent. These patients include all AMIs without sustained complications and include cases with transient sinus tachycardia, atrial fibrillation or hypertension. AMI with pericarditis is also in this class. An example of a class 2 patient is one with an uncomplicated inferior AMI, hypertension and a history of previous AMI.

Class 3 includes patients with single or multiple system involvement. The stage of the disease in this class is acute or chronic, or chronic with exacerbation and/or progressive. There are *single or multiple severe complications with residual impairment.* The response to therapy is incomplete, usually delayed, temporary and intermittent. This class includes patients with persistent CHF, atrio-ventricular disassociation, transient second and third-degree heart block, papillary muscle dysfunction, extension or recurrent AMI, or successfully treated ventricular fibrillation. If a patient has an AMI and also has CHF, COPD, and diabetes, then he/she is in class 3.

Class 4 includes patients with single or multiple system involvement. In this class, the stage of the disease is advanced. There are single or multiple severe complications with residual impairment, and the *response to therapy is minimal or brief or there is no response.* This class includes patients with shock, pump failure, cardiac arrest or respiratory failure. An example of a class 4 patient is one with an AMI, CHF, ventricular tachycardia, cardiac arrest, hypertension, anemia and renal failure.

Specific procedures do not place a patient in one class, except to the extent that the procedures are associated with a diagnosis, e.g., implanted pulsation balloon or use of a respirator.

Results

In Table 1, the sample size, average LOS and coefficient of variation for LOS data in two Baltimore city hospitals for the year 1978 are shown. The same results are also included for CHARGES. We compare DRG 121 (AMI) with the initial refinements of AMI cases suggested by our cardiology panel (IHD-AMI and ARRY-AMI). The average LOS and coefficient of variation in the data grouped as DRG 121 (AMI as primary diagnosis only) are somewhat similar to the average LOS and coefficient of variation in the data set IHD-AMI. However, the average LOS is about two thirds as large in the data set ARRY-AMI; the sample sizes (n) are quite small, however.

Table 1 illustrates the usual methods (and refinements of these methods) currently used by regulatory agencies. From a bigger table with more hospitals, we selected these two hospitals with quite different LOS and charges (a community hospital and a medical center) to study with our more detailed methods. What follows are detailed analyses of DRG 21, the new MD-DATO groups for AMI and the severity groups within these two hospitals.

First, we examine the three methods of grouping. In the DRG system, the MDC AMI is exactly one DRG (because of the difficulty in breaking it into homogeneous subgroups), and it utilizes discharge abstract data. In the MD-DATO system, AMI is divided into four subgroups (based on secondary diagnoses described above) by computer. However, in our current study, we have identified these subgroups from a chart review to avoid any possible coding errors and, hence, to put the method in its most favorable light. In the severity system, the AMI patients are divided into four severity classes by chart review.

Charge Analyses

In Table 2, the charge results for both the medical center and the community hospital are shown. The data for all patients are split into the DRG group, the MD-DATO groups and the AS-SCORE groups. The top half of the table shows the sample size, the average charge, the coefficient of variation and the reduction in variance for all patients within the respective groups. The bottom half of the table shows the same results, but with the deaths and miscoded

TABLE 1. LOS and CHARGES for Two Baltimore City Hospitals
Number of Cases in MDC (N), Number of Cases with AMI (n), Average LOS ($\bar{X}$)
and Coefficient of Variation (CV) by Hospital and AMI Type, 1978

Hospital	AMI			IHD with AMI				Arrythmia with AMI				All AMI		
	N = n	$\bar{X}$	CV	N	n	$\bar{X}$	CV	N	n	$\bar{X}$	CV	n	$\bar{X}$	CV
					LOS									
A. community hospital 1	203	17.5	47.2	207	2	12.0	11.8	138	—	—	—	205	17.4	47.3
B. medical center 2	238	13.6	66.3	938	16	11.8	72.3	246	6	11.7	52.2	260	13.4	66.4
					CHARGES									
A. community hospital 1	203	4149	58.7	207	2	2512	8.7	138	—	—	—	205	4133	58.8
B. medical center 2	238	5480	70.3	938	16	5536	83.4	246	6	4537	48.4	260	5462	70.8

cases omitted from the respective groups. The miscodes were cases that were coded with principal diagnosis AMI by the medical records department, but for which no corroborating evidence for AMI (any two of enzyme changes, typical chest pain or EKG changes) could be found when the charts were reviewed for the patient-severity classification. We also omitted all patients who left against medical advice and all patients who were transferred to another hospital.

First, examine the medical center data for each method of grouping patients. The results are based on 74 randomly selected patients during the period from January to June 1979 with AMI as the principal diagnosis. The average charge in this sample, \$5339, is close to that in Table 1, \$5480 (based on all 1978 cases).

The coefficients of variation for the MD-DATO groups tend to be only slightly smaller than that of the DRG. However, the coefficients of variation of the severity groups are about half the size of those for the DRG and MD-DATO groups. This is true for all data, as well as for the data with deaths and miscodes omitted. The smaller coefficients of variation indicate that the severity groups have better homogeneity with respect to charges per case. One exception is MD-DATO class A (shock) and severity class 4. Both of these groups contain very severely ill patients, many of whom die, and whose charges vary greatly; furthermore, the sample sizes in these groups are small.

There is no reduction in variance using the DRG, since there is only one group. There is only a 4 per cent reduction of variance using the MD-DATO groups for all the data and a 14 per cent reduction in variance using the severity groups for the same data. For the data with deaths and miscodes omitted, the corresponding reductions are 25 per cent for the MD-DATO groups and 63 per cent for the severity groups. The larger reduction in variance indicates that the severity groups are more homogeneous with respect to charges per case.

Separate analysis of variance (ANOVA) tests were computed for each of the grouping systems to determine if the mean values in the subgroups were significantly different. For the MD-DATO groups, the results were not significant ($F_{3,70} = 0.92$ for all data and $F_{3,48} = 57$ with deaths and miscodes omitted). The AS-SCORE means were significantly different ($F_{2,71} = 5.59$, $p < 0.015$ for all data and $F_{2,50} = 26.96$, $p < 0.001$ with deaths and miscodes omitted).

Next, we examine the community hospital data based on 75 randomly selected patients from the first half of 1979 with AMI as the principal diagnosis. Comparing our 1979 sample with data from all of 1978, we see that the average charge for our sample (\$3961) is not much different from the average charge for all cases in 1978 (\$4149). Here again, the severity groups show smaller coefficients of variation and larger reduction in variance. Separate ANOVA tests show that both the MD-DATO means and the AS-SCORE means are significantly different ($p \leq 0.05$).

If we now compare across each row of Table 2 (within each classification), we see that the coefficients of variation are almost always lower in the community hospital, i.e., the cases are more homogeneous than they are in the medical center. It is interesting to note, however, that the *differences* in coefficients of variation between the two hospitals are smallest for the severity groups; indeed, they are almost the same when deaths and miscodes are omitted.

For reduction in variance, the per cent reduction is the same for both hospitals using the patient-severity groups on all data (14 per cent), but it is more than 50 per cent greater in the medical center (63 per cent) than in the community hospital (39 per cent) for the data with deaths and miscodes omitted.

Within each final group, the average charge per case is almost always lower at the community hospital, but again the

TABLE 2. AMI—CHARGES

	All Data					
	Medical Center			Community Hospital		
	n	Average	CV	n	Average	CV
DRG-AMI	74	5339	0.77	75	3961	0.46
	Reduction in Variance = 0			Reduction in Variance = 0		
MD-DATO						
A-SHOCK	12	5071	0.98	4	3713	0.61
B-CHF	8	3733	0.60	34	4604	0.48
C-ARRYTH	26	4989	0.69	22	3845	0.34
D-OTHER	28	6239	0.75	15	2904	0.24
	Reduction in Variance = 0.04 $F_{3,70} = 0.92$, NS			Reduction in Variance = 0.07 $F_{3,71} = 3.43$, $p = 0.05$		
AS-SCORE						
2	31	3486	0.30	36	3254	0.23
3	23	6349	0.49	27	4653	0.38
4	20	7239	0.91	12	4730	0.63
	Reduction in Variance = 0.14 $F_{2,71} = 5.59$, $p < 0.015$			Reduction in Variance = 0.14 $F_{2,72} = 6.53$, $p = 0.005$		

	Deaths and Miscodes Omitted						
	Medical Center				Community Hospital		
	n	Average	CV	Mann-Whitney p = value	n	Average	CV
DRG-AMI	53	5729	0.67		61	3958	0.41
	Reduction in Variance = 0				Reduction in Variance = 0		
MD-DATO							
A-SHOCK	1	8785	0		1	5629	0
B-CHF	6	4516	0.45		24	4660	0.42
C-ARRYTH	22	5497	0.63		22	3845	0.34
D-OTHER	24	6118	0.64		14	2991	0.21
	Reduction in Variance = 0.25 $F_{3,48} = 0.57$, NS				Reduction in Variance = 0.18 $F_{3,57} = 4.09$, $p = 0.03$		
AS-SCORE							
2	28	3694	0.27	$p = 0.12$	36	3254	0.28
3	20	7058	0.38	$p = 0.01$	24	4920	0.35
4	5	11812	0.52	$p = 0.77$	1	8707	0
	Reduction in Variance = 0.63 $F_{2,50} = 26.96$, $p < 0.001$				Reduction in Variance = 0.39 $F_{2,58} = 19.07$, $p < 0.001$		

magnitude of these differences is much smaller within the severity groups. We tested the significance of the differences in charges between the two hospitals within each of the severity groups with deaths and miscodes omitted. Using the Mann-Whitney nonparametric two-sample test, we found the two-tailed probability of the difference in charges in class 2 patients to be $p = 0.12$, in class 3 to be $p = 0.01$, and in class 4 to be $p = 0.77$. We are investigating the reasons for these differences. There is no significant difference ($p > 0.05$) for the least severe and most severe AMI patients, but there is a significant difference between the community hospital and medi-

TABLE 3. AMI—LOS

	All Data					
	Medical Center			Community Hospital		
	n	Average	CV	n	Average	CV
DRG-AMI	74	11.82	0.65	75	14.93	0.43
	Reduction in Variance = 0			Reduction in Variance = 0		
MD-DATO						
A-SHOCK	12	6.83	1.03	4	7.25	1.05
B-CHF	8	9.63	0.68	34	16.32	0.44
C-ARRYTH	26	12.00	0.55	22	16.23	0.33
D-OTHER	28	14.43	0.57	15	11.93	0.27
	Reduction in Variance = 0.12 $F_{3,70} = 0.74$, NS			Reduction in Variance = 0.14 $F_{3,71} = 4.24$, $p < 0.02$		
AS-SCORE						
2	31	9.75	0.29	36	13.61	0.26
3	23	14.39	0.45	27	18.96	0.34
4	20	12.26	1.02	12	9.83	0.87
	Reduction in Variance = 0.03 $F_{2,71} = 2.24$, NS			Reduction in Variance = 0.26 $F_{2,72} = 17.58$, $p < 0.001$		

	Deaths and Miscodes Omitted						
	Medical Center				Community Hospital		
	n	Average	CV	Mann-Whitney p = Value	n	Average	CV
DRG-AMI	53	14.09	0.48		61	16.38	0.34
	Reduction in Variance = 0				Reduction in Variance = 0		
MD-DATO							
A-SHOCK	1	20.00	0		1	18.00	0
B-CHF	6	12.33	0.41		24	18.75	0.33
C-ARRYTH	22	13.09	0.51		22	16.23	0.33
D-OTHER	24	15.21	0.48		14	12.43	0.22
	Reduction in Variance = 0.01 $F_{3,48} = 0.69$, NS				Reduction in Variance = 0.19 $F_{3,57} = 4.38$, $p = 0.02$		
AS-SCORE							
2	28	10.39	0.21	$p = 0.001$	36	13.61	0.26
3	20	16.00	0.33	$p = 0.01$	24	20.13	0.29
4	5	27.40	0.34	$p = 0.50$	1	26.00	0
	Reduction in Variance = 0.60 $F_{2,49} = 33.21$, $p < 0.001$				Reduction in Variance = 0.37 $F_{2,58} = 17.05$, $p < 0.001$		

cal center for patients of moderate severity. Since severity of illness has been controlled for, we can now assess differences in use of technology in laboratory and radiology for these patients in the two institutions, in an effort to determine why class 3 patients are significantly more costly at the medical center, but class 2 and class 4 patients are not.

Length-of-Stay Analyses

In Table 3, the same breakdown of the AMI data is shown, but the variable LOS is now examined. Here, we again find that:

1. The average LOS in our sample of 1979 medical center patients (11.8) is about two days less than for all AMI patients in 1978 (13.6 in Table 1). The average LOS in our sample of community hospital patients is 14.9 days, versus 17.5 days for all of 1978. These changes undoubtedly represent changes in practice patterns for AMIs. It is interesting that we find a decrease in LOS from 1978 to 1979, but Table 2, as compared with Table 1, showed almost no change in charges for the same period.

2. The coefficients of variation tend to be smallest in the severity index (AS-SCORE) groups for each hospital.

3. The community hospital tends to have smaller coefficients of variation than the medical center within each row of the table.

4. The reduction in variance tends to be larger for the severity groups in each hospital.

5. The medical center tends to have a lower median and lower average LOS than the community hospital within each group, and these differences are significant at $p \leq 0.01$ for class 2 and class 3 patients with deaths and miscodes omitted.

6. ANOVA tests show significant differences in means of the AS-SCORE subgroups with deaths and miscodes omitted for the medical center, and all means different for the community hospital.

Table 4 shows how the severity and the MD-DATO groups relate to each other. MD-DATO class A (shock) contains mostly AS-SCORE class 4 patients, but within each of the remaining MD-DATO groups there is wide variation in severity. Thus, these two different methods of grouping patients are not equivalent. The groups based on secondary diagnoses tend to contain patients in all severity levels. One exception, however, is MD-DATO group D (Other) in the community hospital. Almost all of its patients are class 2, as compared with the medical center, which has patients of all three severity levels in MD-DATO group D. This correspondence accounts for the low coefficients of variation for group D patients in the community hospital in Tables 2 and 3.

Table 5 shows that the death rate in our sample at the medical center is 20 per cent, as compared with 13 per cent at the community hospital. (Using the total 1978 discharge abstract data, it was 16 per cent and 10 per cent, respectively.) However, when we take severity into account for the sample data, we see that the death rate at the medical center is actually lower, 75 per

TABLE 4. Relationship Between MD-DATO and AS-SCORE

		Medical Center		Community Hospital	
MD-DATO	AS-SCORE Class	Number of Cases	Average Severity	Number of Cases	Average Severity
A					
	2	0	3.92	0	3.75
	3	1		1	
	4	11		3	
B					
	2	3	2.75	6	3.06
	3	4		20	
	4	1		8	
C					
	2	15	2.50	16	2.27
	3	9		6	
	4	2		0	
D					
	2	13	2.75	14	2.13
	3	9		0	
	4	6		1	

TABLE 5. Death Rates

	Medical Center		Community Hospital	
	n	Death Rate	n	Death Rate
DRG–AMI	74	15/74 = 20%	75	10/75 = 13%
MD-DATO				
A–SHOCK	12	11/12 = 92%	4	3/4 = 75%
B–CHF	8	1/8 = 13%	34	6/34 = 18%
C–ARRYTH	26	0/26 = 0%	22	0/22 = 0%
D–OTHER	28	3/28 = 11%	15	1/15 = 7%
AS-SCORE				
2	31	0/31 = 0%	36	0/36 = 0%
3	23	0/23 = 0%	27	0/27 = 0%
4	20	15/20 = 75%	12	10/12 = 83%

Severity Comparisons

	Medical Center	Community Hospital
	n = 74	n = 75
AS-SCORE		
2	42%	48%
3	31%	36%
4	27%	16%
—		
Average Severity	2.85	2.68

cent in class 4 patients compared with 83 per cent in class 4 patients in the community hospital. Although the difference in these death rates is not significant with such small sample sizes, it does show that the crude death rate alone can be misleading; one needs to take severity of illness into account.

We examined the mortality cases separately to determine in what ways these patients differed in their charges and LOS. The results are shown in Table 6 and are summarized as follows:

1. The coefficients of variation are much larger at the medical center than at the community hospital for both charges and LOS. Apparently, there is more variability in treatment at the medical center for those patients who die.

2. Although the mean charge is larger at the medical center, the median charge is smaller, and the difference between the two institutions, using the Mann-Whitney test, is not significant ($p = 0.62$). The mean charge is larger because the medical center had a few very high charge deaths.

3. The median and mean LOS are less at the medical center, but the difference is not significant ($p = 0.42$).

Discussion and Summary

We have examined in detail three different methods for grouping patients and for adjusting inpatient LOS, charge and death rate data for AMI patients.

The first method utilizes the DRG method and is based entirely on the discharge abstract. This method was found to be the least adequate for grouping AMI patients when compared by two statistical criteria of homogeneity (coefficient of variation and reduction in variance). The DRG method is clearly inadequate for this disease condition. The entire DRG system is now being redefined using the new ICDA-CM-9 data base on LOS from the Commission on Professional and Hospital Activities.

The second method (MD-DATO) groups cases using secondary diagnoses.

TABLE 6. Mortality Cases of AMI

	Medical Center		Community Hospital	
	Charges	LOS	Charges	LOS
N	15	15	10	10
Median	3317	3	3808	6
Mean	5276	6.47	4050	7.1
CV	1.15	1.27	0.69	0.86

The assumption is that secondary diagnoses will add sufficient information to the principal diagnosis of AMI that the homogeneity of the groupings will be statistically improved. Five cardiologists and one internist established subgroupings for AMI based on the secondary diagnoses of shock, congestive heart failure and/or severe arrhythmia, hypertension and/or other minor arrhythmias, and other secondary diagnoses. Subgrouping AMIs by these secondary diagnoses slightly improved the group homogeneity (Tables 2 and 3) as seen by smaller coefficients of variation and larger reductions in variance. Further analyses using new data sets and refinements of the MD-DATO grouping method for discharge abstract data show the same slight improvements in homogeneity. This suggests that for AMI cases, discharge abstract data manipulation may have little potential for producing CHARGE and LOS-homogeneous groups.

The third method uses an implicit four-level measure of patient severity. The statistical criteria using this method are markedly better than those for either the DRGs or the groupings based on physician diagnoses (MD-DATO). A substantial reduction in variance and much smaller coefficients of variation are achieved. Thus, patient groupings that account for severity are essential for appropriate comparisons of one hospital with another for utilization review, reimbursement or quality studies. A simple measure of patient severity adds greatly to an understanding of the data.

The comparison of the medical center and community hospital data showed that the medical center hospital had a substantially more variable patient population with respect to CHARGES and LOS. These patients were treated in a shorter time period at the medical center (Table 3), and their least complicated and most complicated costs were approximately the same (Table 2) as those of patients at the community hospital. After adjustment for severity, the death rate at the medical center hospital changed from higher than that at the community hospital to lower than that at the community hospital, although not significantly lower.

Recognizing that the descriptions of the four AS-SCORE severity classes do not give much guidance to raters, we are now undertaking further research to refine the system. Although the AS-SCORE system has been successful, we are developing a severity index that is even more precisely defined and, therefore, more teachable and more reliable.

Many different agencies need to know how to handle patient discharge data properly. Our study examined three different methods: the discharge abstract, the total chart to get accurate secondary diagnoses and a structured implicit severity measure. Slee[11] has suggested that additional data need to be incorporated into discharge abstracts so that they can be more useful in making decisions about medical care. Our study indicates that a measure of patient severity can be very useful in measuring patient characteristics and hospital productivity.

Acknowledgments

The authors wish to thank George C. Roveti, M.D., F.A.C.P., and Sharon L. Kreitzer, R. N., of Church Hospital, Baltimore, Maryland, for scoring the severity levels for the charts in this study. They also wish to thank Sara May Gold for help with the statistical computations and Susan Stern for manuscript preparation.

References

1. Luft HS, Bunker JP, Enthoven A. Should operations be regionalized. N Engl J Med 1979;301:1364.

2. Fetter RB, Shin Y, Freeman JL, et al. Case mix definition by diagnosis-related groups. Med Care 1980;(Supplement 2):1.

3. Lave JR, Lave LB. The extent of role differentiation among hospitals. Health Serv Res 1971;6:15.

4. Commission on Professional and Hospital Activities. Hospital Mortality, PAS Hospitals, U.SA. 1972–73. Ann Arbor, MI: December 1975.

5. Young WY, Swinkola RB, Hutton MA. Assessment of the AUTOGRP patient classification system. Med Care 1980;18:228.

6. Horn SD, Schumacher DN. An analysis of case mix complexity using information theory and diagnostic related grouping. Med Care 1979;17:382.

7. Horn SD. Validity, reliability, and implications of an index of inpatient severity of illness. Med Care 1981;19:354.

8. Uniform Hospital Discharge Data–Minimum Data Set. Report of National Committee on Vital and Health Statistics, April 1980. (DHEW Publication (PHS)80-1157.)

9. Roveti GC, Horn SD, Kreitzer SL. "AS-SCORE": a multi-attribute clinical index of illness severity. QRB 1980;6:25.

10. Horn SD, Roveti GC, Kreitzer SL. Length of stay variations: a focused review. QRB 1980;6:6.

11. Slee VM. The international classification of diseases: 9th revision (ICD-9). Ann Intern Med 1978;88:424.

MEDICAL CARE
May 1982, Vol. XX, No. 5

Access to Postpartum Sterilization in Southeast Brazil

BARBARA JANOWITZ, PH.D.,* JAMES E. HIGGINS, PH.D.,* DEBORAH C. CLOPTON, M.S.,*
MILTON S. NAKAMURA, M.D.,† AND MICHAEL L. BROWN, B.S.‡

All women hospitalized for delivery over a ten-week period at the largest maternity hospital in Campinas in the State of São Paulo, Brazil, were questioned about their interest in and plans for sterilization. Results from a categorical data analysis indicate that among the study variables, cesarean delivery was the necessary condition for postpartum sterilization and was significantly associated with the patient's ability to pay for services. Further, the variability in the proportion of women sterilized postpartum was almost perfectly explained by a linear model with main effects for parity and for the patient's ability to pay for services.

BRAZIL HAS the largest population in Latin America, with early 1980 census estimates of between 115 and 118 million.[1] Fertility began to decline around 1965, and recent surveys indicate a continuing decrease. The population growth rate was 2.8 for the period between 1965 and 1970, and was estimated to be 2.4 for the period from 1975 to 1979.[2]

During the early 1970s, the government was opposed to family planning, but by 1977 it had modified its position to the extent of promising to provide contraceptives to women for whom pregnancy would create a severe health risk; however, this pledge has not yet been implemented.[3]

One of the country's most important methods of family planning is female sterilization. Data from a recent contraceptive survey indicate that it is the second most prevalent method in São Paulo, the country's richest and most populous state.[4] Of currently married women between the ages of 15 and 44, 64 per cent were using some contraceptive method; of these, 25 per cent had been sterilized. Seventy-five per cent of all sterilizations were performed for women in the year of their last delivery.

Although data on the percentage of sterilizations in conjunction with cesarean delivery are unavailable for the state of São Paulo, surveys conducted in the summer of 1980 reveal that in two Northeast states (Rio Grande do Norte and Pernambuco) 60 per cent of the sterilizations were carried out at the time of cesarean delivery (unpublished data).

Couples able to afford modern methods have many options, so it is not surprising to find that the use of oral contraceptives and condoms is positively correlated with in-

---

From the International Fertility Research Program,* Research Triangle Park, North Carolina; Pontifical Catholic University of Campinas,† Campinas, Brazil; and State University of Campinas,‡ Campinas, Brazil.

Partial support for this work was provided by the International Fertility Research Program, with funds from the U.S. Agency for International Development.

Reprint requests: Barbara Janowitz, International Fertility Research Program, Research Triangle Park, NC 27709.

0025-7090/82/0500/0526/$00.95 © J.B. Lippincott Co.

come[4] or that among women who do not want more children, the percentage sterilized rises with family income.[5]

Twenty years ago, sterilization was performed in Brazil only if medically indicated,[6] but, today, social and economic factors contribute greatly to its justification. Although the Brazilian Medical Ethics Code condemns sterilization, it is permitted in "exceptional" cases when certified by three physicians. However, the rationale for exceptions is not explicit, and many physicians perform the procedure by interpreting the Code to include women who would be at high risk if they become pregnant.

Women who have had previous cesarean deliveries are at high risk, and when they undergo this surgery for the third time, they are almost always sterilized. It may be more difficult for a woman who has never had a cesarean to qualify for a sterilization unless she can meet the other criteria of the "exceptional" cases.

Although the Brazilian Penal Code does not refer to male or female sterilization,[6,7] these procedures are not permitted by the Ministry of Health within its Program of Maternal–Child Health. This attitude is reflected in the national health insurance program, which does not reimburse physicians for sterilization services. Moreover, since the largest employer/employee health insurance plan relies on these government schedules to determine rates of compensation for physician services, it, too, omits payment for surgical sterilization.

As a consequence of government policy, insurance coverage includes the cost of delivery and, usually, a cesarean delivery, but not the concurrent sterilization. Because of the additional personnel, medications, anesthetics and longer hospitalization required, reimbursement rates are necessarily higher for cesarean than for vaginal deliveries, and more generous under private sector plans than through the national insurance system.

Information gathered in four hospitals in the city of Campinas indicates that access to sterilization among women hospitalized for obstetric delivery depends on their socioeconomic status.[8] Although the data collection instrument used was not specifically designed to study access, it may be inferred from these data that sterilization is not equally available to all women.

To gain more information concerning access to sterilization, a study was undertaken at a large maternity hospital in Campinas, chosen because it serves patients of widely varied socioeconomic status and, consequently, of different abilities to pay for surgery.

Data and Method

The study was conducted over a ten-week period, from December 1979 through February 1980. An interview was scheduled after delivery but before discharge for women giving birth during this time. They were asked a number of questions: Did they want any more children? Had they heard about sterilization, and, if so, had they planned to be sterilized? If they had planned to be sterilized, was the procedure actually carried out? If not, why not? And were they considering alternative plans for sterilization? Women who said they were not interested in this method of contraception were asked the reason(s) for their attitude.

During the study period, 2,279 women gave birth, and interviews were completed for 2,194 of them (96.3 per cent). Some information was obtained for 72 of the other cases, but none for the remaining 13.§

The potential market for sterilization services consists of women who want no more children. However, not all women who fall into this group are likely to want to make an irreversible decision to limit their fertility. In an effort to confine the analysis to those who might be expected to have completed their families and not be unduly influenced by the circumstances of pregnancy and delivery, the population analyzed has been limited to the 927 married women who have at least two children, including the current delivery, and are between the ages of 20 and 44 (henceforth described as women who desire no more children).

---

§ Payment status and type of delivery account for the major differences between the women for whom an interview was and was not completed.

The analysis proceeds in steps, the first of which is descriptive, concerning distributions of age, parity (after delivery), monthly family income, education, type of payment for medical services and method of delivery among the 927 women who desire no more children. However, all of them will not necessarily choose sterilization. Therefore, in the second stage of the analysis, attention is directed toward describing how the population desiring no more children chooses sterilization as the preferred method of family planning. The last phase of the analysis is concentrated on the subset of women preferring sterilization services. In this third stage, frequent use is made of methods of estimation and hypothesis testing from cross-classified data.

Results and Analysis

The Market for Sterilization

The second column of Table 1 presents data on the characteristics of women who desire no more children. More than 25 per cent of the women are from 20 to 24 years of age, and more than 60 per cent are under 30; almost 30 per cent have had exactly two children; more than 40 per cent have not completed primary school (less than four years), and less than 20 per cent have eight or more years of education. More than 50 per cent have a monthly family income under $300, and for less than 20 per cent it is at least $500; income is unknown for 14 per cent. Close to 38 per cent of the women had cesarean deliveries; the others, except for one whose type was unrecorded, had vaginal deliveries.

Almost 10 per cent are private patients, 7 per cent are covered by employer/employee–financed plans (*convenio*), more than 70 per cent by government insurance, and 20 per cent have no medical coverage. All women are classified as private patients if they were assigned to the private section of the hospital, and included in this group are some insured patients. It is assumed that women assigned to the "private" section of the hospital will be treated equally, despite the differences in type of payment for services. This inference is borne out by the hospital policy requiring the physician's permission to interview a woman assigned to a private room, but not patients in other accommodations.

Convenio includes all women who pay for their care through a privately financed insurance plan but do not stay in a private room. "Government insured" encompasses the group paying for care with public insurance. "Indigent" includes all women without any medical insurance.

The higher the woman's level of education, the more likely she is to be a private or a *convenio* patient. Similarly, the higher her family's income, the more apt she is to be a private patient. However, 14 per cent of the women did not know or did not wish to report their family income. Because of this large unknown proportion, and since payment status is closely related to income, greater attention will be paid to the impact of the other characteristics (age, parity, education, payment status and type of delivery) on access to sterilization.

Information Concerning Sterilization

To obtain a sterilization, a woman must first have heard of the procedure. Column 3 of Table 1 shows that almost nine out of ten women who want no more children have heard about sterilization. Further, except for parity, the proportion having heard of sterilization increases monotonically with increasing levels of the characteristics in column 1 of Table 1. Indigent patients are least likely (79 per cent) and private patients are most likely (98 per cent) to have heard of it. Women who had cesarean deliveries are more likely to have heard of sterilization (95 per cent) than those giving birth vaginally (86 per cent).

Among the patient characteristics of interest, payment status has the largest spread in the percentage of patients who have heard of sterilization. A sterilized friend is the most important source of knowledge (63 per cent), and a physician is the second (16 per cent). However, private patients are more likely to have received this information from a doctor (25 per cent) than are *convenio* (16 per cent), nationally insured (15 per cent) or nonpaying patients (18 per cent).

Plans for Sterilization

Of all women who want no more children and have heard of sterilization, as

TABLE 1. Desire for No More Children, Knowledge of and Plans for Sterilization of Married Women Aged 20–44, With at Least Two Children, by Selected Characteristics

Characteristic	A Number Desiring No More Children	Have Heard of S*: Per Cent of A	Have Heard of S*: B Number of	S* Plans: Per Cent of B
All characteristics	927	89.2	827	58.2
Age				
20–24	263	83.6	220	36.4
25–29	310	91.0	282	58.5
30–34	208	90.9	189	76.7
35–39	115	93.0	107	70.1
40–44	31	93.5	29	55.2
Parity				
2	276	89.5	247	30.0
3	291	90.4	263	65.4
≥4	360	88.1	317	74.1
Monthly family income				
$ 0–$199	285	84.2	240	58.3
$ 200–$299	185	87.0	161	57.1
$ 300–$499	163	93.9	153	57.5
≥$500	163	98.8	161	68.3
Unknown	131	85.5	112	45.5
Education (years)				
0–3	381	82.7	315	61.0
4	240	92.1	221	57.9
5–7	131	91.6	120	54.2
≥8	175	97.2	171	56.1
Payment status				
Indigent	94	78.7	74	40.5
Government insured	662	89.1	590	58.1
Convenio	66	93.9	62	61.3
Private	92	97.8	90	72.2
Unknown	13	84.6	11	45.5
Delivery type				
Vaginal	576	85.5	494	44.9
Cesarean	350	94.8	332	78.0
Unknown	1	100.0	1	—

* S = Sterilization.

shown in column 5 of Table 1, 58 per cent said that they planned to be sterilized. The percentage of women who plan to be sterilized is curvilinear with age, with 36 per cent of women between the ages of 20 and 24 planning to be sterilized, as compared with 77 per cent between the ages of 30 and 34 and 55 per cent from 40 to 44 years of age. As parity increases, so does the proportion planning sterilization. Among women at parity 2, 30 per cent plan to be sterilized, as compared with 65 per cent at parity 3 and 74 per cent who have given birth to four or more children.

Education and family income up to $500 monthly show little association with plans for sterilization. However, 10 per cent more women at the highest monthly income level plan to be sterilized than those less privileged.

Private patients were more likely to say they planned to be sterilized than insured patients, and insured more than the indigent. Finally, women having cesarean deliveries were almost twice as likely to be planning sterilization as those who had vaginal deliveries.

Information Concerning Sterilization Facilities

To be sterilized postpartum, a woman must know that services are offered at the hospital where her baby is born. Column 3 of Table 2 shows that 78 per cent of women who want no more children and who have heard about sterilization know that it is available at the hospital. This proportion is lowest for the age group between 20 and 24 years and highest for women 30 years of age and over, an expected relationship since many under the age of 30 may be less strongly committed to the decision to be sterilized than older mothers and, therefore, have sought less information about the procedure. Knowledge shows little variation with parity, but is surprisingly low for women at parity 4 and over, compared with women at lower parities. Knowledge concerning services increases with income and education, and only 60 per cent of women who had vaginal deliveries were as likely to be as knowledgeable as those who had cesarean deliveries. The relationship to payment status is particularly strong. More than 95 per cent of private patients know that sterilization is offered at the hospital, as compared with only 43 per cent of nonpaying patients.

The varied answers to the questions concerning availability may be a function not only of the woman's knowledge of the existence of the hospital's sterilization facilities, but also of her perception that these services are being offered to her. It may be that some women, particularly those from the lower socioeconomic groups, answer negatively because they believe that the facilities are not available to them, rather than that the hospital does not actually possess such facilities.

Who Gets Sterilized?

The last two columns on Table 2 display, respectively, the numbers and percentages of women sterilized among those planning it who also knew that services were available at the hospital. Of these 376 women, 226 (60 per cent) received sterilizations. The proportion sterilized increases with income, education and for age groups up to those between 35 and 39 years of age. For parity (including the current birth), the percentage sterilized reaches a maximum (74 per cent) at parity 3. Payment status has an important influence on who is sterilized. More that 80 per cent of the private patients who planned sterilization were sterilized, as compared with 68 per cent of the *convenio*, 54 per cent of the government insured and 46 per cent of the indigent patients. Finally, women with cesarean deliveries were 16 times more likely to receive postpartum sterilization than those who had vaginal deliveries.

To investigate further the influence of payment status on who is sterilized, atten-

TABLE 2. Knowledge About Availability of Sterilization and Attainment of Postpartum Sterilization for Selected Characteristics of Married Women Aged 20–44 Who Have at Least Two Children, Desire No More Children and Are Planning Sterilization

Characteristic	A Number of S* Plans	Know S* Offered at Hospital		Postpartum S*	
		Per Cent of A	B Number of	Per Cent of B	Number
All characteristics	481	78.2	376	60.1	226
Age					
20–24	80	66.3	53	54.7	29
25–29	165	77.6	128	56.3	72
30–34	145	84.1	122	63.1	77
35–39	75	80.0	60	66.7	40
40–44	16	81.3	13	61.5	8
Parity					
2	74	81.1	60	48.3	29
3	172	81.4	140	73.6	103
≥4	235	74.9	176	53.4	94
Monthly family income					
$ 0–$199	140	70.7	99	47.5	47
$ 200–$299	92	78.3	72	44.4	32
$ 300–$499	88	83.0	73	63.0	46
≥$500	110	89.1	98	76.5	75
Unknown	51	66.7	34	76.5	26
Education (years)					
0–3	192	68.8	132	47.7	63
4	128	78.1	100	61.0	61
5–7	65	87.7	57	68.4	39
≥8	96	90.6	87	72.4	63
Payment status					
Indigent	30	43.3	13	46.2	6
Government insured	342	76.7	263	54.0	142
Convenio	38	97.4	37	67.6	25
Private	65	95.4	62	83.9	52
Unknown	5	20.0	1	100.0	1
Delivery type					
Vaginal	222	57.2	127	5.5	7
Cesarean	259	96.1	249	87.9	219
Unknown	—	—	—	—	—

* S = Sterilization.

tion will be restricted to the 375 women who were both planning sterilization and aware of its availability at the hospital (225, i.e., 60 per cent were sterilized).[‖]

The variables of interest in this analysis are postpartum sterilization, payment status, age, education, parity and type of delivery, each classified at two levels. Payment status is defined as public (indigent and government insured) and nonpublic (*convenio* and private). Age is dichotomized by year groups, 20–29 and 30–44, and education by year groups, zero–four and at least five. Because of the practice of performing sterilizations for women having their third cesarean delivery, parity (including present birth) is

---

[‖] There are 375 women available for this analysis, instead of 376 as shown at the top of column 4 of Table 2, because one woman could not be classified by payment status.

dichotomized by two and at least three, and the method of delivery is categorized as vaginal or cesarean.

The strongest first-order association between the explanatory variables and sterilization is exhibited by the method of delivery (Pearson $\chi^2 = 236.9$, 1DF, $p < < 0.01$). Among women who had cesarean deliveries, 88 per cent were sterilized, compared with 6 per cent of the others. In descending rank of importance of first-order association with sterilization, the explanatory variables following method of delivery are payment status, education, parity and age. However, the association between age and sterilization does not achieve significance at $\alpha = 0.05$.

In general, to be sterilized postpartum a woman must have an abdominal delivery. For patients in this category, the proportion sterilized (0.88) is the same regardless of payment status. However, a significantly greater proportion of women in the non-public payment subpopulation (0.83 versus 0.60) had cesarean births (Pearson $\chi^2 = 18.8$, 1DF, $p < 0.01$). Non public payment patients who had vaginal deliveries receive proportionally more sterilizations (0.25 versus 0.03) than public payment patients (Pearson $\chi^2 = 13.4$, 1DF, $p < 0.01$).

Cesarean delivery is, in general, the necessary condition for postpartum sterilization and, as shown above, is significantly associated with payment status. Consequently, it is of interest to focus on the association between sterilization and each of the explanatory variables (age, parity and education), while controlling for payment status. When payment status is controlled using the Mantel-Haenszel procedure,[9] only parity remains significantly associated with sterilization (Mantel-Haenszel $\chi^2 = 8.2$, 1DF, $p < 0.01$).

A close look at the four subpopulations defined by the levels of payment status and parity (Table 3) reveals that the respective proportions of women sterilized can be represented almost perfectly by an additive model with main effects for payment status and parity (i.e., there is no interaction between payment status and parity). The methods of Grizzle, Starmer and Koch[10] are applied to estimate and test the statistical significance of the main effects. Table 4 displays estimates of the payment status and parity effects, along with test statistics for their significance. Both main effects are highly significant. The estimates indicate that, regardless of parity, the difference in the proportions sterilized between women whose payment is non-public and public is 26 per cent. Similarily, the difference between the per cent of women of parity 3 or more and of parity 2 who were sterilized is 18 per cent, regardless of payment-status level.

Why Women Are Not Sterilized

Of the 375 women who planned to be sterilized and knew about available ser-

TABLE 3. Prevalence of Sterilization by Payment Status and Parity (for the 375 Women Who Were Both Planning Sterilization and Aware That It Was Available at the Hospital)

		Sterilized		
		Yes		No
Payment Status	Parity	Number	(Proportion)	Number
Public	2	16	(0.38)	26
	≥3	132	(0.56)	102
Nonpublic	2	16	(0.64)	9
	≥3	61	(0.82)	13

TABLE 4. Analysis of Variance Table for a Main Effect Model of the Proportion of Women Sterilized (for the 375 Women Who Were Both Planning Sterilization and Aware That It Was Available at the Hospital)

Source of Variation	Estimate of Effect	DF	χ^2 Test Statistic
Payment status	0.26	1	27.0*
Parity	0.18	1	8.1*

* Denotes significance at $\alpha = 0.01$.

vices, 40 per cent were not sterilized postpartum. The most frequent reason given was type of delivery (38 per cent); of the women who had vaginal deliveries, 48 per cent (57/120) gave this as a reason for failure to be sterilized. Other frequent explanations for not having undergone sterilizations were: a) it was not customary for the doctor to do them, or he did not want to operate (29 per cent); b) the physician considered that the woman was too young or that she did not have enough children (15 per cent); and c) the woman either did not discuss the possibility of, or actually arrange for, sterilization with her doctor (7 per cent).

Conclusions and Policy Implications

Data from this study show that the method of payment for care substantially influences the eventuality of the sterilization. Wealthier women who are private or *convenio* patients are more likely to be sterilized than are the poor, whose care is financed through government insurance or who are indigent. Results also show that women sterilized postpartum almost always have had cesarean deliveries. Regardless of whether she is sterilized, a cesarean delivery is more likely as the socioeconomic status of the woman rises.

What then can be done to increase access to this method of family planning among poor women in Brazil? A more liberal interpretation of the Medical Ethics Code that would consider women to be at high risk for reasons other than those associated with previous cesarean birth would improve access to postpartum sterilization. As a consequence of current policies, women with high family income who have a history of cesarean deliveries may qualify for sterilization on medical grounds, whereas poor women who have had vaginal deliveries are turned down for sterilization because of the very nature of their delivery.

Physicians who do perform postpartum sterilizations cannot charge for their services under the national or private insurance plans. Other countries have increased access to sterilization by modifying medical insurance plans (thus making it more attractive for surgeons to perform the operation) or by paying physicians directly. In the United States, as in many other countries, most medical insurance plans cover both tubal ligations and vasectomies, and the extension of coverage to include sterilization has doubtless increased the number carried out.

Brazil is well on its way to modifying its posture regarding family planning, and there are signs that a publicly supported program will soon be undertaken. The Minister of Health, Waldir Arcoverde, recently announced official adoption of a family planning program throughout the country that will include free access to tubal ligations.

This article has demonstrated that access to sterilization in Brazil is differentially affected by legal and institutional constraints on delivery of services. These results may be generalized to other medical services and geographic locations. The present controversy in the United States over family planning and abortion could give rise to conditions similar to those in Brazil.

Acknowledgment

The authors wish to thank Irene Rosenfeld for her assistance.

References

1. Censo revela que O Brasil ainda não tem 120 milhões. Jornal Do Brasil, October 10, 1980.

2. Rowe PM. Demographic Profiles: Brazil. Bureau of the Census, U.S. Department of Commerce, January 1980.

3. Merrick TW. Brazil's population to be hemisphere's largest in 21st century. Intercom 1979;7:11.

4. Nakamura MS, Morris L, Janowitz B, et al. Contraceptive use and fertility levels in São Paulo State, Brazil. Stud Fam Plann 1980;11:236.

5. Janowitz B, Anderson JE, Morris L, et al. Service availability and the unmet need for contraceptive and sterilization services in São Paulo State, Brazil. Int Fam Plann Perspect 1980;6:10.

6. Nakamura MS, Gonçalves MEF, de Carvalho MA. The status of surgical contraception in Brazil. In: Surgical Methods of Family Planning: The Role of the Private Physician. Research Triangle Park, NC: International Fertility Research Program, 1981.

7. Da Fonseca RM. The law and the condition of women in Brazil. In: Law and the Status of Women: An International Symposium. Center for Social Development and Humanitarian Affairs, United Nations. New York: Columbia University School of Law, 1977:11.

8. Janowitz B, Lewis J, Clopton D, Nakamura MS. Post-Partum Sterilization in Sao Paulo State, Brazil. J Biosoc Sci (in press).

9. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719.

10. Grizzle JE, Starmer FC, Koch GG. Analysis of categorical data by linear models. Biometrics 1969;25:489.

BEST AVAILABLE COPY

C5170-1
7/28/82

Medical Care

June 1982
Vol. 20 No. 6
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Outcome-Based Doctor–Patient Interaction Analysis: I. Comparison of Techniques. II. Identifying Effective Provider and Patient Behavior

Psychosocial Factors Affecting Adherence to Medical Regimens in a Group of Hemodialysis Patients

Knowledge, Attitude, and Compliance Dimensions of Antibiotic Therapy With PPIs: A Community Pharmacy-Based Study

A Survey of Patients' Receipt of Prescription Drug Information

Satisfaction With Ambulatory Care and Compliance in Older Patients

Participants in Noncompliance Research: Compliant or Noncompliant?

Health Diaries: Strategies for Compliance and Relation to Other Measures

Patient Characteristics That Influence Physician Behavior

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. BERKI, M.A.
University of Michigan

ROBERT H. BROOK, M.D., SC.D.
University of California
Los Angeles

GORDON H. DE FRIESE, PH.D.
University of North Carolina

EUGENE FEINGOLD, PH.D.
University of Michigan

BRANT E. FRIES, PH.D.
Yale University

THOMAS R. GODKINS, B.S., P.A.
University of Oklahoma

HAROLD HUNTER, DR. P.H.
Washington, D.C.

SHERMAN A. JAMES, PH.D.
University of North Carolina

ANTHONY L. KOMAROFF, M.D.
Brigham and Women's
Hospital, Boston

JAMES P. LOGERFO, M.D., M.P.H.
University of Washington

NANCY MILIO, PH.D.
University of North Carolina

DUNCAN NEUHAUSER, PH.D.
Case Western Reserve University

JOHN E. WARE, JR., PH.D.
The Rand Corporation

ALONZO S. YERBY, M.D.
Harvard University

ALFRED YANKAUER, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. **Address for subscription information, orders or changes of address:** (except Japan) 2350 Virginia Avenue, Hagerstown, MD 21740. In Japan, contact Woodbell Scope, 11-11, Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. **Annual subscription rates:** U.S. $45.00 individual, $50.00 institution; all other countries except Japan, $55.00 individual, $60.00 institution; special student rate, $27.00. Single copies $6.00. Copies will be replaced without charge if the publisher receives a request within 60 days of the mailing date in the U.S. or within 5 months in all other countries. Printed in U.S.A. © Copyright 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, MD, and at additional mailing offices.
POSTMASTER: Send address changes to Medical Care, 2350 Virginia Avenue, Hagerstown, MD 21740.

June 1982
Volume 20, Number 6

CONTENTS

ORIGINAL ARTICLES

Outcome-Based Doctor–Patient Interaction Analysis: I. Comparison of Techniques. *Thomas S. Inui and William B. Carter with Walter A. Kukull and Virginia H. Haigh* 535

Outcome-Based Doctor–Patient Interaction Analysis: II. Identifying Effective Provider and Patient Behavior. *William B. Carter and Thomas S. Inui with Walter A. Kukull and Virginia H. Haigh* 550

Psychosocial Factors Affecting Adherence to Medical Regimens in a Group of Hemodialysis Patients. *K. Michael Cummings, Marshall H. Becker, John P. Kirscht, and Nathan W. Levin* 567

Knowledge, Attitude, and Compliance Dimensions of Antibiotic Therapy With PPIs: A Community Pharmacy-Based Study. *Audrey R. Gotsch and Salvatore Liguori* 581

A Survey of Patients' Receipt of Prescription Drug Information. *Louis A. Morris* 596

Satisfaction With Ambulatory Care and Compliance in Older Patients. *Margaret W. Linn, Bernard S. Linn, and Shayna R. Stein* 606

Participants in Noncompliance Research: Compliant or Noncompliant? *Janice A. Ramsay* 615

Health Diaries: Strategies for Compliance and Relation to Other Measures. *Geoffrey R. Norman, Allan H. McFarlane, David L. Streiner, and K. Neale* 623

Patient Characteristics That Influence Physician Behavior. *Elizabeth M. Hooper, Loretto M. Comstock, Jean M. Goodwin, and James S. Goodwin* 630

COMMUNICATIONS

Do Postcard Reminders Improve Influenza Vaccination Compliance? A Prospective Trial of Different Postcard "Cues." *Eric B. Larson, James Bergman, Fred Heidrich, Barbara L. Alvin, and Ronald Schneeweiss* 639

Validation of a Patient Satisfaction Scale: Theory, Methods and Practice. *Barbara S. Hulka and Stephen J. Zyzanski* 649

BOOK REVIEWS

Gentle Vengeance: An Account of the First Year at Harvard Medical School, by Charles LeBaron. *Steven Jonas* 654

Health Planning and Regulation: The Decision-Making Process, by Drew Altman, Richard Greene, and Harvey M. Sapolsky. *Steven Sieverts* 656

LETTERS TO THE EDITOR 659

MEDICAL CARE
June 1982, Vol. XX, No. 6

Psychosocial Factors Affecting Adherence to Medical Regimens in a Group of Hemodialysis Patients

K. MICHAEL CUMMINGS, PH.D., M.P.H., MARSHALL H. BECKER, PH.D., M.P.H., JOHN P. KIRSCHT, PH.D., AND NATHAN W. LEVIN, M.D.

The present research was designed to identify psychosocial correlates of adherence among patients receiving hemodialysis at two outpatient clinics. The 116 participants were interviewed concerning their knowledge of treatment, health beliefs, treatment history, social support, personal characteristics, and adherence to the medication, diet and fluid-intake aspects of the regimen. In addition, medical record data were obtained on serum phosphorus and serum potassium levels, and on between-dialysis weight gains. The magnitude of relationships between predictors and adherence measures varied, depending on the method used to measure adherence. For the self-report measures, beliefs concerning the efficacy of the behavior and barriers to the behavior, along with reported family problems, proved to be consistent predictors. Other beliefs and characteristics did not contribute significantly. For the medical chart information, however, the predictive factors were less consistent. In general, situational factors seem to be the major contributors to patient adherence, and adherence itself is seen as a complex and multidimensional phenomenon.

FOR PATIENTS with end-stage renal disease, replacement of aspects of the different functions of the kidney is obtained through hemodialysis, dietary management and medication. Hemodialysis removes toxic fluids and metabolic end products from the blood stream by use of an artificial kidney machine. Unfortunately, such treatments alone are usually not sufficient to prevent or alleviate the symptoms that result from kidney failure, and most patients are also required to follow a complex and trying diet regimen, which restricts protein, sodium, potassium and fluid intake. The objective of dietary management is to assist hemodialysis treatments by reducing the amounts of urea, creatinine and electrolytes (especially potassium, sodium and phosphate) that must be excreted.[1,2] Abuse of dietary restrictions can lead to a buildup of toxic fluids and metabolic end products in the blood stream, which may, in turn, result in nausea, weakness, cardiovascular problems and even death. In addition to dietary

---

From the Department of Cancer Control and Epidemiology, Roswell Park Memorial Institute, Buffalo, New York; the School of Public Health, University of Michigan, Ann Arbor, Michigan; and the Henry Ford Hospital, Detroit, Michigan.

This research was supported by a grant from the Michigan Department of Public Health, Division of Chronic Disease Control, Lansing, Michigan.

Requests for reprints: K. Michael Cummings, Ph.D., M.P.H., Department of Cancer Control and Epidemiology, Roswell Park Memorial Institute, 666 Elm Street, Buffalo, NY 14263.

0025-7079/82/0505/0567/$01.20 © J. B. Lippincott Co.

controls, most patients are requested to take phosphate-binding medications, and many patients are also given vitamin supplements, drugs to reduce blood pressure, drugs to prevent blood clotting and other medications as well.[2] Thus, successful treatment of end-stage renal disease depends to a large degree upon the patient's willingness to cooperate with a therapeutic plan that is often very restrictive and difficult to follow.

However, despite the adverse consequences associated with poor compliance, there is substantial evidence of considerable abuse of dietary and medication regimens by hemodialysis patients.[3–16] An averaging of compliance rates from many studies[10,12,13,15,16] reveals that about one third of all patients fail to comply with both between-dialysis fluid limitations (measured by between-dialysis weight gains) and diet restrictions (measured by serum potassium levels), and about half do not follow instructions for taking phosphate-binding medication (measured by serum phosphorus levels).

Determinants of Compliance

Relatively few of the studies reported in the literature have attempted to identify factors associated with compliance behaviors among hemodialysis patients. Furthermore, a common aim of many of these studies was to identify variables that could be used as criteria to select patients for a limited number of treatment slots; as a result, much of this research is descriptive, focusing mainly on characteristics of the patient (e.g., age, sex, socioeconomic status, intelligence, psychologic functioning) and upon features of the patient's social situation (e.g., family size, quality of home environment), variables that are relatively enduring and immutable.

A seemingly more promising approach to understanding patient compliance behaviors has concentrated on a set of health belief variables. This Health Belief Model (HBM) seeks to explain the way in which an individual chooses among alternative health actions (i.e., the decision to comply or not comply); it postulates that an individual's decision to take action relative to a particular health threat is a function of his/her beliefs along the following subjective dimensions: 1) beliefs about the nature of the health threat, in terms of personal susceptibility to that danger and its perceived severity if it should occur; 2) beliefs that specific actions, such as staying on a diet or taking medications as prescribed, will prevent or reduce the threat; and 3) beliefs about the barriers or costs that may be associated with taking the recommended action. Compliance is seen as more likely in patients for whom, in the presence of a perceived threat, the recommended action is viewed as beneficial and possible at a tolerable cost. It is important to note that the emphasis is on what the individual believes, rather than on what is held to be scientifically correct.

There is considerable research evidence that supports the usefulness of the HBM in accounting for variation in the extent to which people follow preventive practices,[17] deal with symptoms[18] and respond to illness.[19] In a recent study by Hartman and Becker,[15] the HBM was also used to predict dietary regimen adherence in a group of 50 hemodialysis patients; their findings indicated that, compared with noncompliers, compliers were more likely to: 1) view a list of sequelae commonly associated with noncompliance as having potential serious impact on life; 2) view the treatment as beneficial; and 3) report fewer difficulties in following the treatment regimen.

Besides health beliefs, other variables have been demonstrated to influence extent of compliance. Since recommendations for treatment vary in their demands on the patient (e.g., with regard to complexity and extent of required behavior change), a number of studies have evaluated the characteristics of the regimen in relation to compliance.[20,21] In general, the more the patient is told to do, the greater is the likelihood of noncompliance. Several studies have shown higher rates of medication errors (both overdosing and underdosing) when many doses of a medication are prescribed. The amount of change required by the health recommendation also seems to be related to compliance. In an investigation involving cardiac patients, Davis[22] noted that patients tended to make those changes that required the least amount of adjustment for their lifestyle; modifications of personal habits, such as diet and alcohol consumption, were the least well followed.

Some knowledge of treatment instructions and of the reasons for the treatment also appear as important conditions for attaining acceptable levels of patient cooperation. Davis[23] found that when the significance of the regimen was not conveyed to patients, compliance suffered. In a study involving over 600 hemodialysis patients, Borkman[11] reported a significant association between patients' understanding of dietary restrictions and staff-rated dietary compliance. It is quite possible that the complexity of the regimen is related to compliance because of demands made on the patient's ability to comprehend, retain and apply details of the regimen.

Several studies have noted the important influence of support given by family members.[12,15,16] Hartman and Becker[15] found that hemodialysis patients who reported fewer family problems and received more assistance from their spouses (both in general and in health-related matters) complied better with a dietary regimen. In a study of 40 home hemodialysis patients, Pentecost et al.[12] reported that patients' compliance with their daily fluid restrictions was positively correlated with the quality of communication among family members. As a negative factor, social isolation has been found to be related to noncompliance.[16]

A number of studies (including research on hemodialysis patients) point to the importance of patient–provider relationships. Kaplan-DeNour and Czaczkes[24] found an association between the expectations and evaluations of patients by dialysis clinic staff and subsequent patient compliance behaviors. Wertzel[26] reported a similar finding, showing that hemodialysis patients tended to conform to providers' expectations about their behavior.

The present investigation examines the relationships of health beliefs, knowledge about the treatment regimen, regimen complexity, family and provider support, and personal characteristics to compliance in a group of ambulatory hemodialysis patients. Adherence to three components of the treatment plan was examined: 1) taking phosphate-binding medicine; 2) following dietary restrictions (especially the limiting of foods high in potassium content); and 3) limiting fluid intake. Findings from this research effort are used to suggest ways for improving patient compliance.

Method

Setting and Subjects

The study was conducted in two outpatient dialysis clinics located in southeastern Michigan. Patients were accepted into the study if they had no unusual physical or mental disabilities that could be expected to interfere with their participation in an interview, were between 18 and 80 years of age, and had been receiving dialysis treatments for a minimum of three months. The eligible study group consisted of 120 patients (72 per cent of the total patient population from both clinics), each of whom received a letter briefly describing the research and asking him/her to participate. Interviews were conducted by trained interviewers while the patients were receiving their dialysis treatments at the clinics.

The interview included questions concerning the patient's treatment history (e.g., length of time receiving dialysis treatment, eligibility for kidney transplant); health beliefs; knowledge of the treatment regimen; perceptions about the amount of support received from family members, friends and clinic medical staff; personal characteristics; and self-reported compliance to components of the treatment regimen. Of the 120 potential study participants, only four refused to be interviewed (n = 116, 97 per cent cooperation rate).

Characteristics of patients at the two clinics were compared to determine the appropriateness of pooling data from both sites. Two-sample t-tests and chi-square analyses were performed on sociodemographic, medical history, health beliefs, and knowledge and social support variables, as well as on compliance levels. The analyses indicated no statistically significant difference at the $p < 0.10$ level between patients from the clinic sites. Thus, all subsequent analyses are based on combined data from the two clinics.

Patients ranged in age from 21 to 76 years (mean 54.8). Of the initial 116 study patients, 54 per cent were male, and half were white. The medial level of formal education was high school graduation. Median family income was between $10,000 and $10,999 per year, with 16 per cent receiving less than $5,000 annually. Most of the participants (67 per cent) were married, with 9 per cent never married and the remainder widowed, divorced or separated. Eight patients (7 per cent) reported that they lived alone. Fifteen per cent of the sample indicated that they were currently employed either full-time or part-time. Of those individuals not currently employed, 72 per cent indicated that they had stopped working because of their illness. The average length of time the study patients had been receiving dialysis treatments was 29 months; most patients (87 per cent) were receiving dialysis treatments three times per week. Only a quarter of the patients (n = 28) indicated that they expected to have a kidney transplant in the future, and 14 patients (12 per cent) reported having received a kidney transplant.

Measures

Patient Compliance

Compliance with taking phosphate-binding medicine, with following dietary restrictions and with instructions for limiting fluid intake was measured using two methods of assessment: 1) information obtained from the patient's medical chart (i.e., blood chemistries and weight gains between dialysis treatment); and 2) patient self-report.

Medical Chart Information. Dialysis patients need to control the amount of phosphorus in their blood, since their kidneys can no longer perform this function adequately. Phosphate-binding medications are used to help control the amount of phosphorus that is absorbed in the intestinal tract. Thus, the serum phosphorus level (SPHL) is an indirect measure of patient compliance with taking phosphate-binding medicine. In this study, compliance with taking phosphate-binding medicine was assessed by averaging patients' SPHLs over a period of two to four months. In all, four measures were obtained, three taken before and one taken after the interview. For the great majority of patients, SPHLs were obtained routinely once every month.

Since patients with renal failure are unable to excrete potassium adequately, dietary restrictions are imposed to help the patient maintain proper levels of potassium in the blood. Thus, serum potassium level (SPL) is an indirect indicator of compliance with dietary restrictions of foods high in potassium (primarily fruits and vegetables). In this study, compliance with dietary restrictions was evaluated by averaging patients' SPLs over a period of two to three weeks. Six measures were obtained, three before and three after the interview. SPLs were assessed routinely each time the patient came for dialysis.

To avoid excess fluid buildups, dialysis patients are restricted in the amount of fluid permitted between dialysis treatments. The degree of fluid restriction depends on patients' ability to excrete fluid on their own. Thus, interdialysis weight gain is an indirect measure of patient compliance with fluid restrictions. In this study, compliance with limiting fluid intake was appraised by averaging the patient's daily weight gains (DWGs) over a period of two to three weeks. The DWG was calculated by dividing the patient's between-dialysis weight gain by the number of days since the last dialysis treatment. Between-dialysis weight gains were calculated by subtracting from each patient's predialysis weight the previous treatment's postdialysis weight. Predialysis and post dialysis weights were obtained routinely at each treatment session.

As measures of compliance, the medical chart assessments have the advantage of being relatively unaffected by human judgments; however, these measures may not be accurate indicators of compliance since they can be affected by factors unrelated to patient compliance behavior. For example, both serum potassium and serum phosphorus levels can be influenced by the presence of a co-existing catabolic process or by the degree of adequacy of the dialysis treatment. Between-dialysis weight gains can be affected by patients' activity level, the number of calories in their diet and by the amount they perspire.

Patient Self-Reports. As part of the interview designed to assess health beliefs and knowledge about their illness, patients were asked to rate the degree to which they usually comply with instructions about their phosphate-binding medicine, their diet and their between-dialysis fluid limit. Patients rated the direction and degree of their compliance with each of the three components of the medical regimen on a 7-point rating scale, where responses ranged from "poor" to "excellent" compliance.

While fairly direct, patient reports of behaviors are subject to several sources of potential invalidity, for example, the natural desire to report "good" behaviors and the difficulty of recalling instances of noncompliance. In general, studies that have compared self-reported compliance with other methods of measuring regimen compliance (e.g., physiologic measures such as blood samples and urine tests to detect trace levels of medication, pill counts, health outcomes) have found that patients tend to overestimate their compliance with recommendations.[26,27]

Health Beliefs

Items pertaining to particular health belief dimensions were combined into indices on the basis of content and interitem relationships. The major indices included: 1) perceived susceptibility to the sequelae of *non*compliance with taking: phosphate-binding medicine (e.g., "What chance do you think there is that you would develop bone disease?"); dietary restrictions (e.g., "What chance do you think there is that you would have weak and tired muscles?"); and fluid intake limits (e.g., "What chance do you think there is that you would have shortness of breath?"); 2) perceived severity of sequelae associated with *non*compliance with: taking phosphate-binding medicine (e.g., "How serious do you think it would be if you were to get high levels of phosphorus in your blood?"); dietary restrictions (e.g., "How serious do you think it would be if you were to get high levels of potassium in your blood?"); and fluid intake limits (e.g., "How serious do you think it would be if you were to store up too much fluid in your body between dialysis treatments?"); 3) beliefs about the likely benefits of: taking phosphate-binding medicine (e.g., "How much do you think the phosphorus medicine helps to prevent you from developing bone disease?"); the diet (e.g., "How much do you think the diet helps you to prevent you from becoming weak and tired?"); and the fluid limit (e.g., "How much do you think limiting the amount of fluid you drink prevents you from having shortness of breath?"); 4) perceived barriers to: taking the phosphate-binding medicine as prescribed (e.g., "Did you ever miss taking your phosphorus medicine because the medicine was costing you a lot of money?"); following the diet (e.g., "Did you ever not follow the diet because you don't always have time to prepare special meals?"); and limiting fluid intake (e.g., "Did you ever not stick to your fluid limit because you felt the fluid limit was too strict?").

The susceptibility, severity and benefit items were measured using 7-point rating scales. Indices were constructed by computing mean scores. Estimates of internal consistency for these indices ranged from 0.57 to 0.92, with the indices of perceived severity and perceived benefits yielding the lowest and highest levels of consistency, respectively. The "barriers" items were coded either "yes" or "no"; these indices were constructed by adding the number of "yes" responses to items that assess factors interfering with a patient's compliance to a particular regimen (i.e., seven items concerning barriers to compliance with taking phosphate-binding medicine and to following dietary restrictions, and five items involving barriers to compliance with fluid restrictions).

Knowledge of Regimen Purpose

The patient's knowledge of how each of the regimen's three components is supposed to help was gauged by having two nurses (one from each clinic) rate each patient's response to three open-ended questions: 1) "How does the phosphorus medicine help you?" 2) "How does the diet help you?" and 3) "How does the fluid limit help you?" The nurses' ratings were scaled from "1" (does not understand purpose of treatment regimen) to "7" (understands exactly). The nurse raters were blinded with regard to which responses corresponded to which patients. The interjudge reliabilities for the three knowledge measures ranged from 0.65 to 0.70.

Knowledge of Treatment Instructions

To assess understanding of the limits imposed by the special diet, the dietitian in each clinic rated the patients' responses to the question: "What types of foods are restricted in your diet?" The dietitian's rating was scaled from "1" (not knowledgeable about types of foods restricted) to "7" (very knowledgeable). In judging a patient's response, the dietitians were instructed to consider both the accuracy and completeness of the answers.

Dietitians in each clinic were also asked to score their patients' responses to the question : "What is your daily fluid limit?" ("1" = correct, "2" = incorrect, which included "don't know"). Patients' knowledge of treatment instructions for taking phosphate-binding medicine was not assessed.

Regimen Complexity

A measure of regimen complexity relative to the taking of phosphate-binding medicine was constructed by multiplying the patient's report of the frequency of taking the medicine by the reported dosage of the medicine taken each time. Regimen complexity was not assessed for the diet and fluid intake restrictions.

Support From Family and Friends

A three-item index measured the extent to which the patients perceived their spouse, other family members and friends as making it easier for them to do what they are supposed to do for their kidney disease. Items employed 7-point rating scales, and the index was constructed by computing the mean for the three items (internal consistency for the index was 0.65).

Support From Medical Staff

A two-item index measured the degree to which the patients viewed their doctors and other members of the dialysis staff as making it easier for them to follow their regimens. Again, items used 7-point rating scales, and the index was the mean for the two items (internal consistency for the index was 0.95).

Family Problems

A five-item index was constructed to represent the effects of the patient's illness on family life. Items included questions con-concerning the extent to which patients viewed their illness as: 1) interfering with the amount of time they spent with their spouse; 2) causing problems with their marriage; 3) creating financial problems for their family; 4) interfering with their role as a parent; and 5) causing other problems in their family life. All responses were obtained along 7-point rating scales, and an index was constructed by computing the mean for the five items (internal consistency for the index was 0.82).

Table 1 presents the means, ranges, and standard deviations for the compliance

TABLE 1. Means, Ranges, Standard Deviations and Number of Respondents for Compliance Measures and Independent Variables

Measure	N	Range	Mean	Standard Deviation
Compliance				
Phosphorus level (mg/dl)	101	2.40–10.1	5.84	1.34
Potassium level (mEq/L)	111	3.50–6.20	4.85	0.55
Daily weight gain (kilograms)	111	0.33–2.61	1.34	0.48
Patient's report of compliance with taking phosphate-binding medicine	98	1–7	6.02	1.21
Patient's report of compliance with dietary restrictions	111	1–7	4.88	1.46
Patient's report of compliance with limiting fluid intake	100	1–7	5.00	1.41
Independent Variables				
Perceived susceptibility to sequelae of noncompliance with taking phosphate-binding medicine	101	2–7	5.50	1.10
Perceived susceptibility to sequelae of noncompliance with dietary restrictions	111	3–7	5.64	0.92
Perceived susceptibility to sequelae of noncompliance with limiting fluid intake	108	3–7	5.97	0.92
Perceived severity of sequelae of noncompliance with taking phosphate-binding medicine	101	2–7	5.92	1.20
Perceived severity of sequelae of noncompliance with dietary restrictions	111	2–7	5.61	1.08

TABLE 1. Continued

Measure	N	Range	Mean	Standard Deviation
Perceived severity of sequelae of noncompliance with limiting fluid intake	108	4–7	6.12	0.84
Beliefs about benefits of taking phosphate-binding medicine	98	2–7	6.20	1.01
Beliefs about the benefits of the diet	111	3–7	5.81	1.04
Beliefs about the benefits of limiting fluid intake	108	3–7	6.00	0.88
Barriers associated with taking phosphate-binding medicine	101	0–6	1.93	1.57
Barriers associated with following dietary restrictions	111	0–7	2.82	1.68
Barriers associated with limiting fluid intake	108	0–5	1.74	1.28
Knowledge of how the phosphate-binding medicine helps	98	1–7	2.58	1.74
Knowledge of how the diet helps	111	1–6	2.80	1.39
Knowledge of how the fluid limit helps	102	1–6	3.51	1.27
Knowledge of foods restricted in the diet	111	1–7	4.51	1.32
Knowledge of daily fluid intake	108	1–2	1.62	0.49
Complexity of the phosphate-binding medicine regimen	96	1–16	5.45	3.63
Support from family and friends	114	1–7	5.52	1.56
Support from the medical staff	115	1–7	6.34	1.11
Family problems	93	1–7	5.12	1.75

measures and for each of the independent variables.

Results

Using the criterion compliance level set by the medical staff (an average SPHL of less than 5.5 mg/dl), only 30 per cent of the patients were found to be taking the phosphate-binding medicine appropriately. However, based on a compliance criterion level of an average SPL of less than 5.5 mEq/L, 86 percent of the patients appeared to be following their dietary restrictions. Finally, using an average weight gain of less than 3.0 kg between dialysis treatments as a criterion, 59 per cent of the patients could be classified as compliant in limiting their fluid intake.

According to the patient report measures of adherence, patients are most compliant with following instructions for taking their phosphate-binding medicine and least compliant with staying on their diet. The medical chart and patient report measures of compliance were weakly correlated (compliance with taking phosphate-binding medicine, r = 0.36; compliance with dietary restrictions, r = 0.17; and compliance with limiting fluid intake, r = 0.06).

Correlates of Compliance

The product moment correlations of health belief measures, knowledge measures, regimen complexity and social support measures with the measures of compliance are presented in Table 2.

The magnitude (and, in some cases, the direction) of the correlations varies depending on the method used to measure compliance. Patients' beliefs about the benefits of the phosphate-binding medicine are significantly correlated with both self-reported compliance with taking the medicine and with SPHL. Beliefs about the benefits of the diet are strongly correlated with the patient's reported compliance with the diet and moderately correlated with the patient's SPL. Beliefs about the benefits of limiting fluid intake are significantly related to self-reported compliance with limiting fluid intake, but are unrelated to the patient's DWG. The number of barriers reported by the patient as interfering with his/her ability to comply with the treatment regimen is significantly and negatively correlated with each of the three patient-reported measures of compliance and with the patient's SPL and DWG (i.e., patients reporting fewer barriers were more compliant). The patient's perception of the *severity* of sequelae that might result from noncompliance with the regimen was not associated with any of the compliance measures; however, perceptions about *susceptibility* to sequelae associated with not following a treatment regimen were positively and moderately correlated with self-reported compliance with dietary and fluid limits.

The patient's knowledge of the purpose of the treatment regimen is significantly related to SPHL but is not associated with self-reported compliance in taking the phosphate-binding medicine. Knowledge of the purpose of the treatment regimen is unrelated to both the patient report and medical chart measures of dietary and fluid limit compliance. It is of general interest to note that 43 per cent of patients taking phosphate-binding medicine reported that they had no idea what the purpose of the medicine was; 24 per cent did not know the purpose of their diet, and 7 per cent did not know how their fluid restriction was supposed to help them.

Patients' knowledge of the types of foods restricted in their diets was not associated with dietary compliance, nor was knowledge about daily fluid limit related to keeping within daily fluid restrictions. Most patients (80 per cent) were fairly knowledgeable (i.e., were rated by the nutritionist at 4 or above on a 7-point knowledge scale) concerning the types of foods restricted in their diet. Eighteen per cent of the patients reported having no idea what their daily fluid limit was, and an additional 19 per cent reported their fluid limit incorrectly.

Complexity of the phosphate-binding medicine regimen (in terms of frequency and dosage schedule) is significantly correlated with the patient's SPHL (i.e., the more complex the regimen, the lower the compliance) but is unrelated to the patient's own reports about taking the medicine.

Measures of support from family members and friends are not related to any of the compliance measures. However, the patients' perception of the effects of their illness on their family life is significantly correlated with each of the three patient report measures of compliance and with the patient's DWG (i.e., the more patients see their illness as disrupting their family life, the poorer their compliance). Support from members of the medical staff at the clinic is moderately associated with the patients' reported compliance with taking their phosphate-binding medicine and staying on their diet.

Table 3 presents correlations of sociodemographic and medical history variables with measures of compliance with the three components of the treatment regimen. Again, the magnitude (and, in some cases, the direction) of the correlations de-

TABLE 2. Correlations (Pearson Product Moment) of Independent Variables and Compliance Measures

	Compliance Measures					
Independent Variables	Phosphorus Level (mg/dl) (SPHL)	Patient's Report of Compliance With Taking Phosphate-Binding Medicine	Potassium Level (mEq/L) (SPL)	Patient's Report of Compliance With Dietary Restrictions	Daily Weight Gain (Kilograms) (DWG)	Patient's Report of Compliance With Limiting Fluid Intake
Health Beliefs						
Perceived susceptibility to sequelae of noncompliance	−0.10 (101)	0.09 (98)	0.05 (111)	0.22* (111)	−0.08 (108)	0.17 (100)
Perceived severity of sequelae of noncompliance	0.05 (101)	0.18 (98)	0.09 (111)	0.09 (111)	0.01 (108)	0.03 (100)
Beliefs in the benefits of the regimen	0.21* (98)	0.37* (98)	0.15 (111)	0.56* (111)	−0.08 (98)	0.31* (100)
Barriers interfering with following the regimen	−0.14 (101)	−0.42* (98)	−0.24* (111)	−0.54* (111)	−0.26* (108)	−0.38* (100)
Knowledge						
Knowledge of the purpose of the regimen	0.22* (98)	0.02 (98)	−0.10 (111)	0.05 (111)	0.02 (99)	−0.07 (100)
Knowledge of treatment instructions	NM	NM	0.08 (111)	0.04 (111)	−0.10 (105)	0.01 (100)
Characteristics of the Regimen						
Regimen complexity	−0.35* (96)	0.01 (96)	NM	NM	NM	NM
Social Support						
Support from family and friends	−0.03 (100)	0.09 (97)	−0.01 (109)	0.10 (109)	−0.01 (106)	0.10 (98)
Support from the medical staff	0.07 (101)	0.20* (98)	0.12 (110)	0.25* (110)	−0.07 (107)	0.10 (99)
Family problems	−0.08 (83)	−0.35* (83)	−0.14 (88)	−0.27* (88)	−0.30* (86)	−0.27* (86)

Notes: N indicated in parentheses; NM = not measured.
* $p < 0.05$.

TABLE 3. Correlations (Pearson Product Moment) of Sociodemographic Variables, Medical History Variables and Compliance Measures

	Compliance Measures					
Independent Variables	Phosphorus Level (mg/dl) (SPHL)	Patient's Report of Compliance With Taking Phosphate-Binding Medicine	Potassium Level (mEq/L) (SPL)	Patient's Report of Compliance With Dietary Restrictions	Daily Weight Gain (Kilograms) (DWG)	Patient's Report of Compliance With Limiting Fluid Intake
Sociodemographic Variables						
Age (older patients are better compliers)	0.22* (101)	0.17 (98)	0.14 (110)	0.27* (110)	0.34* (110)	0.14 (99)
Sex (females are better compliers)	0.03 (101)	0.06 (98)	0.11 (111)	0.02 (111)	0.21* (108)	0.01 (100)
Family income	−0.04 (94)	−0.08 (91)	−0.01 (102)	0.06 (102)	0.06 (99)	0.11 (91)
Level of education	−0.11 (100)	−0.04 (97)	0.02 (110)	−0.08 (110)	−0.04 (107)	−0.06 (99)
Medical History Variables						
Length of time receiving dialysis treatments (the longer the time on dialysis the poorer the compliance levels)	−0.11 (101)	−0.03 (98)	−0.20* (111)	−0.11 (111)	−0.13 (108)	−0.04 (100)
Expect kidney transplant in the future (those who expect a transplant are poorer compliers)	−0.10 (94)	−0.17 (91)	−0.19* (102)	−0.19* (102)	−0.22* (99)	−0.02 (93)
Previous failed kidney transplant (those who have previously had a transplant are poorer compliers)	0.05 (101)	0.13 (98)	0.03 (111)	0.28* (111)	0.20* (108)	0.01 (100)

Note: N indicated in parentheses.
* $p < 0.05$.

pends on the method employed to assess compliance. Among the sociodemographic variables examined, only age is consistently associated with adherence; older patients tended to be somewhat better compliers. Using DWG measure as an indicator of compliance, findings show that females were more compliant with the fluid limit than were males.

With regard to medical history variables, the results reveal a modest, consistent, negative association between the length of time the patient has been receiving dialysis treatments and adherence; patients who have been on dialysis for a longer period of time tended to be poorer compliers. Patients' expectations about receiving a kidney transplant were significantly related to the patient report and medical chart measures of dietary compliance and to the patient's DWG. Patients who indicated that they expected to have a kidney transplant in the future were generally poorer compliers. The data also show failure of a previous kidney transplant to be associated with increased daily weight gains and with lower patient-reported compliance with the diet.

In addition to the bivariate correlations, multiple regressions were run on each dependent variable, utilizing the several psychosocial measures as predictors. For the three *self-report measures* (medication, diet and fluid limits), the results were consistent. In each case, beliefs about the benefits of the actions, beliefs regarding barriers to the actions and perceptions of family problems formed the set of significant predictors. Other measures, such as beliefs in susceptibility and severity, regimen complexity, social support, and so on, did not enter the regressions as significant contributors. The psychosocial factors accounted for between 30 per cent (fluid limit) and 53 per cent (dietary compliance) of the variation in the self-reported measures of adherence.

Medical chart information regarding adherence was much less consistently predicted by the set of psychosocial factors. For SPHL, regimen complexity, perceptions of benefits and beliefs about susceptibility constituted the significant variables. For DWG, perceptions of barriers was a predictor, but so was age. Other measures yielded nonsignificant relationships, and, overall, 22 per cent of the variance was accounted for by the factors. For SPL, no measure yielded significant predictions, and only 13 per cent of the variance could be accounted for by the set of factors.

Discussion

Findings from this study suggest that patient compliance behavior is determined to a large extent by situational factors. For example, the most common barriers reported by patients as interfering with their attempts to stay on their diets included having a craving for foods not allowed in their diet (82 per cent), being away from home (61 per cent) and difficulty in preparing special meals (38 per cent). Difficulties frequently mentioned by patients as barriers in following fluid limit recommendations included getting thirsty (76 per cent) and problems with always having to measure the amount of fluid one can drink (43 per cent). Common problems in taking the phosphate-binding medicine included forgetting (58 per cent), being away from home (54 per cent) and having the prescription for the medicine expire (28 per cent). Complexity of the instructions for taking the phosphate-binding medicine was also associated with noncompliance. The results from the multiple regressions indicated that perceived barriers are significant predictors of most measures of compliance.

The assumption (proposed by HBM) that an individual decides to comply (or not comply) with a treatment regimen based on beliefs about probabilities of disease occurrence, severity of the disease should it occur, and likelihood of effective treatment was generally not supported by these data. Examination of the mean scores on the health belief indices shows that the majority of patients perceived themselves to be susceptible to the sequelae of noncompliance, viewed the consequence of these sequelae as serious, and saw the treatment regimen as effective in preventing these sequelae from occurring. However, what patients *believe* they are doing or not doing is quite consistently related to beliefs regarding the efficacy and costs of the action.

Contrary to expectation, level of knowledge of treatment instructions was not associated with compliance. Perhaps the situational determinants of compliance override any effects that knowledge may have—or possibly the patient's ability to rely on family members to insure that they follow their regimen eliminates the need for exact knowledge of treatment instructions.

Knowledge of the purpose of the treatment regimen was found to be significantly related to compliance with taking phosphate-binding medicine, but unrelated to compliance with following dietary restrictions and with limiting fluid intake. In general, patients were not very knowledgeable about the purpose of their treatment regimen.

In contrast to several studies reported in literature,[12,15,16] findings from this study showed little association between support given by family members and compliance. However, the extent to which patients viewed their illness as disruptive of their family life was related to compliance. Younger patients viewed their illness as causing more problems in family life than did older patients.

Practice Implications

Given this study's results, it is suggested that interventions aimed at improving the compliance behavior of hemodialysis patients should focus on reducing environmental barriers that interfere with the patient's ability to comply. Thus, reducing the complexity of the schedule for taking the phosphate-binding medicine would appear to be one way to increase medication compliance. The use of dose dispensers and/or special reminders (such as stickers to put on the refrigerator) may be effective in helping patients to remember when to take the medicine. Having patients write out a menu for each meal a week or so in advance may be helpful in allowing them to plan some variety in the foods they eat and to help them in deciding what foods they need to purchase. Discussions with patients about the types of foods they can order when they eat in a restaurant may help them to stay on their diet when they are away from home. To help patients stay within their fluid limit, it may be useful to review the environmental cues commonly associated with consumption of fluids (e.g., cued by times of day, television watching, sitting alone, mealtimes) and to develop strategies to manipulate these cues or to discuss alternative ways of reacting to the cues (e.g., eating hard candies to make saliva and help reduce dryness in the mouth). For patients who have difficulty pacing the amount of fluid they are allowed to have between dialysis treatments, one might suggest that the patient keep a container of water large enough to hold the total amount of fluid allowed in the refrigerator; they would either drink only from the container or remove an equivalent amount of water from the jug, thus allowing the patients to keep track of amount of fluid consumed between treatments.

The particular strategies used to improve compliance should always depend on the individual. However, the basic approach suggested involved analyzing the environmental factors likely to interfere with compliance behavior and rearranging the environment to facilitate the occurrence of the desired behavior. Different aspects of compliance are virtually unrelated, however, so that it is necessary to deal with each behavior in its own right. Furthermore, the results suggest that patients form quite definite assessments of how they are doing—judgments that may not correspond to "objective" measures. Practically speaking, it appears that multiple assessments of compliance, using a variety of measurement approaches, are the best way to assure accuracy in gauging levels of compliance.

References

1. Anderson CF, Nelson RA, Margie SD, et al. Nutritional therapy for adults with renal disease. JAMA 1973;223:68.

2. Burton BT. Current concepts of nutrition and diet in diseases of the kidney. J Am Diet Assoc 1974;65:623.

3. Retan JW, Lewis HY. Repeated dialysis of indigent patients for chronic renal failure. Ann Intern Med 1965;64:284.

4. Shea EJ, Bogdan DF, Freeman RB, et al. Hemodialysis for chronic renal failure. IV. Psychological consideration. Ann Intern Med 1965;62:558.

5. Pendras JP, Erickson RV. Hemodialysis: a successful therapy for chronic uremia. Ann Intern Med 1966;64:293.

6. Meldrum MW, Wolfram HD, Rubini ME. The impact of chronic hemodialysis upon the socioeconomics of a veteran patient group. J Chronic Dis 1968;21:37.

7. Abram HS, Moore GL, Westervelt FB Jr. Suicidal behavior in chronic dialysis patients. Am J Psychiatry 1971;127:1119.

8. Kaplan DeNour A, Czaczkes JW. Personality factors in chronic hemodialysis patients causing noncompliance with medical regimen. Psychosom Med 1972;23:333.

9. Hickey KM. Impact of kidney disease on patient, family, and society. Soc Casework 1972;53:391.

10. Neff S, Baez A, Slifkin R, et al. Outpatient dialysis. Arch Intern Med 1973;131:317.

11. Borkman TS. Hemodialysis compliance: the relationship of staff estimates of patients' intelligence and understanding to compliance. Soc Sci Med 1976;10:385.

12. Pentecost RL, Zwerenz B, Manuel J. Intrafamily identity and home dialysis success. Nephron 1976;17:88.

13. Blackburn SL. Dietary compliance of chronic hemodialysis patients. J Am Diet Assoc 1977:70:31.

14. Magrab PR, Papadopoulov ZL. The effect of a token economy on dietary compliance for children on hemodialysis. J Appl Behav Anal 1978;10:573.

15. Hartman PE, Becker MH. Noncompliance with prescribed regimen among chronic hemodialysis patients: a method of prediction and education diagnosis. Dialysis and Transplantation 1978;7:978.

16. Procci WR. Hemodialysis patients and dietary abuse. Psychosomatics 1978;19:16.

17. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974;2:354.

18. Kirscht JP. The health belief model and illness behavior. Health Educ Monogr 1974;2:387.

19. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care 1975;13:10.

20. Blackwell B. Drug therapy: patient compliance. N Engl J Med 1974;289:249.

21. Hulka BS, Cassel JC, Kupper L. Disparities between medications prescribed and consumed among chronic disease patients. In: Lasagna L, ed. Patient Compliance. Mt. Kisco, NY: Futura, 1976;123, 152.

22. Davis MS, Eichhorn R. Compliance with medical regimens: a panel study. J Health Human Behav 1963;4:240.

23. Davis MS. Variations in patients' compliance with doctors' orders: analysis of congruence between survey responses and results of empirical investigations. J Med Educ 1966;41:1037.

24. Kaplan DeNour A, Czaczkes JW. Bias in assessment of patients on chronic dialysis. J Psychosom Res 1974;18:217.

25. Wertzel H, Wollrath P, Ritz C, et al. Analysis of patient–nurse interaction in hemodialysis units. J Psychosom Med 1977;21:359.

26. Gordis L, Markowitz M, Lillienfeld AM. Inaccuracy of using interviews to estimate patient reliability in taking medications at home. Med Care 1969;7:49.

27. Cummings KM, Kirscht JP, Becker MH, et al. Construct validity comparisons of three methods of measuring patient compliance. Health Serv Res 1982 (in press).

MEDICAL CARE
June 1982, Vol. XX, No. 6

Knowledge, Attitude, and Compliance Dimensions of Antibiotic Therapy With PPIs

A Community Pharmacy-Based Study

AUDREY R. GOTSCH, DR.P.H., AND SALVATORE LIGUORI, PH.D.

This pilot project was conducted to measure the effectiveness of patient package inserts (PPIs) when controlled interventions by pharmacists are increased. The PPIs were designed to be easily understood by patients and included information on possible side effects, interactions, compliance and instructions on how to take the medication. PPIs for penicillin and ampicillin were evaluated. Data analysis yielded the following results for respondents receiving PPIs, as compared with those who did not:

- **Satisfaction with the amount of drug information received;**
- **Increase in knowledge concerning the drug's proper use;**
- **Higher rates of compliance;**
- **High rating given to the "helpfulness" of the PPI information;**
- **Interest in receiving PPIs routinely, particularly with new prescriptions.**

The PPIs developed for this study were found to be valuable tools to respondents for understanding how to take their medications, both in terms of knowledge and behavior. Verbal reinforcement of the PPI by the pharmacist was also found to be helpful.

PATIENT PACKAGE INSERTS (PPIs) are standardized monographs of information about a drug or drug class intended for users of prescription medicines. PPIs have provoked much controversy as to their rationale, benefits and drawbacks.[1] Opponents argue that a beneficial role in therapeutics for PPIs has not been demonstrated by adequate testing. Further, they state that PPIs will probably not be read by patients and that, if they are read, PPIs may promote self-diagnosis and prescription drug interchange among patients; may produce suggestion-induced side effects and adverse reactions among susceptible patients; may frighten patients away from needed therapy regimens; and may increase the legal liability of the drug maker, prescriber and dispenser. Proponents argue that PPIs should accompany prescribed drugs dispensed to patients because the safe and effective use of such medications requires that patients be informed of their benefits, risks and directions for use; furthermore, patients have a basic right to such information.

From the College of Medicine and Dentistry of New Jersey—Rutgers Medical School and the the College of Pharmacy, Rutgers University, Piscataway, New Jersey.

Financial support for this study was received from the Hunterdon Health Fund by the Office of Consumer Health Education, Department of Environmental and Community Medicine, College of Medicine and Dentistry of New Jersey—Rutgers Medical School.

0025-7079/82/0505/0581/$01.25 © J. B. Lippincott Co.

The amount of information that may be included creates two categories of PPIs. *Full disclosure* PPIs present complete benefit–risk information and appear particularly useful with potentially hazardous "elective" drugs. A clear discussion of the existence of less hazardous and even less effective alternatives to therapy with such drugs in full-disclosure PPIs could help the patient, in consultation with the prescriber, to make a better educated benefit–risk decision about the course of treatment. *Selective disclosure* PPIs, on the other hand, give the patient basic information specifically aimed at ensuring that the prescribed medication is taken with optimal safety and therapeutic effect.

Green and Faden[2] have estimated that it is highly likely that well-designed PPIs will increase patients' awareness of drug information, aid patients' comprehension and recall of drug information, help patients recognize side effects, and improve patients' confidence and competence in self-care. They have estimated that it is moderately likely that such PPIs will increase patients' interest in drug information, enhance patients' informed consent, reduce patients' errors in dosage and dosage omissions, increase patients' reporting of side effects and decrease the fear of side effects among patients. Enhanced compliance with certain therapy regimens, especially those prescribing antibiotics, through reduction of dosage and omission errors, has also been demonstrated with PPIs.[1]

Methods to increase compliance with prescribed drug regimens are needed. Studies have found varying, but in all cases high, rates of noncompliance with drug therapy regimens. One study reports noncompliance ranging from 20 to 82 per cent[3]; another found noncompliance at 35 per cent.[4] A review of the literature on noncompliance states that it reaches as high as 60 per cent in some groups.[5] Broadly conceived, noncompliance includes omissions of scheduled doses, premature cessation of therapy, administration errors such as instilling suspensions for oral use into the ear, timing errors, overutilization and utilization with medicinal agents not intended by the prescriber.

Providing counseling alone, without written instruction. has several shortcomings.[6] The patient may forget or not even hear the information, due to stress and other factors. The physician may be too busy to provide sufficient information or explanation, or may be unaware of the information the patient needs. The patient may be discouraged from asking questions of the physician or the pharmacist. The prescription may be dispensed to a client other than the patient, with a possible loss of information in the transfer. In addition, the patient may receive the prescription in the pharmacy from a nonpharmacist not trained in counseling patients about prescription drugs.

Yet while counseling alone has its drawbacks, written information alone, without verbal counseling, has also shown a negative effect on compliance.[7] This points to the hypothesis that a combination of written and oral instruction would be a more effective way to convey drug information to patients.

To test this hypothesis, we have adapted a widely used model in health education that measures an educational tool or process according to its effect on patient *knowledge, attitudes* and *behavior*. The thesis underlying this model is that given adequate and understandable information, a patient should become more knowledgeable about his/her medical condition and its treatment. With increased knowledge, a change in attitude would be anticipated, disposing the patient to cooperate with the treatment regimen. Finally, as a result of the positive changes in knowledge and attitudes, the patient is more likely to adopt behaviors that will lead to improved health. In the case of drug therapy regimens, the desired behavior is compliance. Ideally, then, according to the health education model, a well-designed and comprehensible PPI, combined with a planned verbal consultation, should augment patient knowledge, leading to an attitude of cooperation with the physician and pharmacist, which should result in improved adherence to the medication regimen.

Description of the Study

The study was supported in part by a grant from the Hunterdon Health Fund and was conducted by the Office of Consumer Health Education of the Department of Environmental and Community Medicine, Rutgers Medical School, College of Medicine and Dentistry of New Jersey, in collaboration with the Department of Pharmacy Administration, College of Pharmacy, Rutgers University.

This study was designed as a pilot study in preparation for a statewide study that would randomly assign the study population to various educational interventions at different types of pharmacies. As a pilot study, the principal objective of this study was to evaluate the effects of PPIs on the knowledge, attitudes and compliance of patients on a short course of therapy with either penicillin V, penicillin G or ampicillin. The quasi-experimental research design included three sequential levels of educational intervention, requiring an increasing amount of involvement by the pharmacist. The participating pharmacists were trained in their role just before each intervention was to begin in order to control for pharmacist bias toward one intervention. The pharmacists knew that their role in the study would change but did not know until each intervention began the scope of their role for the specific interventions. In addition to the training for each educational intervention, a printed reminder sheet highlighting their role for the specific educational intervention was hung at the dispensing counter.

Potential subjects were identified for each educational intervention at each study site and were not included in more than one educational intervention. The stages were as follows:

Educational Intervention I. The study drugs were dispensed to patients in this group with no planned educational intervention by the pharmacists. At dispensing, the pharmacists responded to any questions patients may have had but transmitted no further written or oral drug information. This intervention constitutes the control group for the study.

Educational Intervention II. The study drugs were dispensed with a PPI to patients during this intervention. The pharmacists responded to any questions patients may have had, as in Intervention I, and they dispensed the PPI as unobtrusively as possible, making no overt reference to it. Patients in Educational Intervention II thus received only the PPI, with no planned verbal reinforcement by the pharmacist.

Educational Intervention I. The study drugs were dispensed to patients in this tients in this group. In addition, patients received verbal counseling by the pharmacists, which included notifying them about the PPI, encouraging them to read it and highlighting its salient points. Thus, patients in Educational Intervention III received both the PPI and verbal consultation by the pharmacist.

Study Sites and Respondents

Two community pharmacies under the same management were selected as sites for the study. Although the pharmacies are both located in central New Jersey, one serves primarily an urban community and the other a more rural community.

Both sites have procedures for providing minimal patient counseling, which expedited the implementation of this study. In a follow-up study, it would be interesting to include sites that do not normally provide patient counseling in order to determine if patient expectations vary according to past pharmacy practices.

Patients who presented with new prescriptions for the study drugs were given a consent form to read while their prescriptions were being prepared. The form explained that a telephone survey was to be conducted by the CMDNJ-Rutgers Medical School and the College of Pharmacy. Information collected would be treated confidentially and would be used for statistical purposes only. Patients could signify their agreement to participate in the survey by signing the forms and providing their telephone numbers. Pharmacists collected the forms from all patients to whom they had been given. Although data were not gathered on patients who refused to sign the consent form, according to the pharmacists, very few patients declined to participate. Care was taken throughout the recruitment procedure to avoid explaining the reasons for the survey or describing the study objectives.

Patients who signed the consent forms were called for the telephone interview not more than three days before their prescribed regimen should have been completed. For each educational intervention, a sample of 62 respondents was obtained before the next educational intervention was implemented. Although it would have strengthened this study to have randomly assigned patients to educational interventions in order to control for seasonal and similar variations, it was not possible since this study made use of only two pharmacy sites. Consequently, in order to control the educational interventions, it was necessary to complete one educational intervention for a group of patients before the next intervention was begun. This was necessary in order to ensure that a patient from Intervention I, for example, did not overhear the information being given to a patient in Intervention III who happened to be receiving the same prescriptions at the same time. Simultaneous collection of data from the three intervention groups could be accomplished in a study with more sites in which the pharmacies and consequently the patients could be randomly assigned to different educational interventions. However, in this study it was not possible to collect data on all three educational intervention groups at the same time.

Respondents who were not available when first called were called again at a later time, but always before the prescribed regimen was due for completion. When data collection began in September 1978, the incidence of prescriptions for the study drugs was rather infrequent as compared with the last stage of data collection in late November and early December. Therefore, it took longer to assemble a sample of 62 respondents in Educational Intervention I than in Intervention II or Intervention III.

Interview Questionnaire

An interview questionnaire was developed with a minor modification that allowed it to be used both for patients who had and who had not received the PPI. The instrument included questions about what information respondents remembered receiving when their prescriptions were dispensed, their satisfaction with the information received and any further information they would have liked to receive. Respondents who reported receiving the PPI were asked if they had read it, if they had found it helpful, if they were satisfied with the amount of information it contained, if it was easy to read and if they would like PPIs dispensed routinely with prescription drugs. Respondents who reported receiving verbal counseling from the pharmacist were also asked if the consultation had been helpful, if they were satisfied with the amount of information presented by the pharmacist and if they would like the pharmacist to counsel them routinely when they have prescriptions dispensed.

Such questions may be taken as measures of attitude toward information received in the three educational interventions.

In an effort to keep the telephone interview time to a minimum, the pharmacy patient profile cards were used to collect data regarding the patient's sex and age. In a follow-up study, additional items such as years of education, family income, occupation and type of illness should be included in order to develop a more complete profile of respondents.

The instrument also asked respondents three questions probing their knowledge of the study drug prescribed. A final section dealt with medication-taking behavior and, as a measure of compliance, asked respondents to count the tablets or capsules, or to estimate the amount of liquid remaining to be taken at the time of interview.

The questionnaire was pretested during the month of August. As a result, it was edited slightly to improve language and the flow pattern of questions.

PPI Development

In determining the basic content to be included in the study PPIs for penicillin and ampicillin, the package inserts for those drugs, intended for health professionals, were used as source material. The most difficult and time-consuming aspects of the PPI design involved deciding what information to include for a *selective disclosure* PPI and simplifying complicated technical language for lay comprehension, while still providing adequate, useful information. A three-section format on one page was developed. The first (introductory) section identifies the drug by its generic and more common brand names, briefly explains the drug's actions and instructs the patient to report certain pre-existing conditions to the physician if he/she has not already done so. The second section contains instructions for taking the medication, and the third section presents advice regarding side effects, storage of the medication and general precautions. The PPIs developed for this study appear in the Appendix.

Results and Discussion

Forty-four per cent of the 186 study participants were males. The percentage of males tended to increase from Interventions I to III (36 per cent were males in Intervention I, 46 per cent in II, and 51 per cent in III). Both children and adults received the study drugs; however, if a patient was under 16 years of age, the parent was contacted for the interview. Nearly twice as many patients received prescriptions for penicillin (66 per cent) as for ampicillin (34 per cent). Intervention I showed a larger number of ampicillin prescriptions (29) than did Intervention II (18) or III (17). All of the prescriptions were for oral dosage forms of the study drugs, with 26 per cent of the patients receiving liquid preparations and 74 per cent receiving capsules or tablets. Therapy regimens of less than ten days were prescribed for 52 per cent of the patients, while ten-day regimens were prescribed for 46 per cent. Only 2 per cent of patients received prescriptions for therapy regimens longer than ten days. The length of regimen did not vary appreciably among the three interventions.

Most respondents accurately described the intervention to which they had been exposed. Fifty-eight per cent of the respondents in Educational Intervention I reported receiving neither written nor verbal instructions when they obtained their prescribed medications. Eighty-two per cent of respondents in Intervention II were aware that they had received a PPI with their antibiotic. In Educational Intervention III, 94 per cent of respondents were aware of receiving a PPI, and 89 per cent reported having received verbal consultation from the pharmacist as well (Table 1).

TABLE 1. Respondents' Awareness and Use of PPI by Type of Educational Intervention

Awareness and Use	Educational Intervention* (Per Cent of Total)		
	II	III	II and III Combined
Aware PPI Received	82 (N = 62)	94 (N = 62)	88 (N = 124)
Read PPI	96 (N = 51)	91 (N = 58)	94 (N = 109)

* Educational Interventions:

I. Study drugs dispensed to patients with no planned educational intervention.

II. Study drugs dispensed to patients with a PPI.

III. Study drugs dispensed to patients with a PPI and reinforced with verbal counseling by the pharmacist.

It should be noted that the pharmacist's verbal reinforcement in Intervention III significantly increased the proportion of respondents who were aware of the PPI over the proportion of respondents aware of the PPI in Intervention II ($Z = 1.97$, $p < 0.05$).

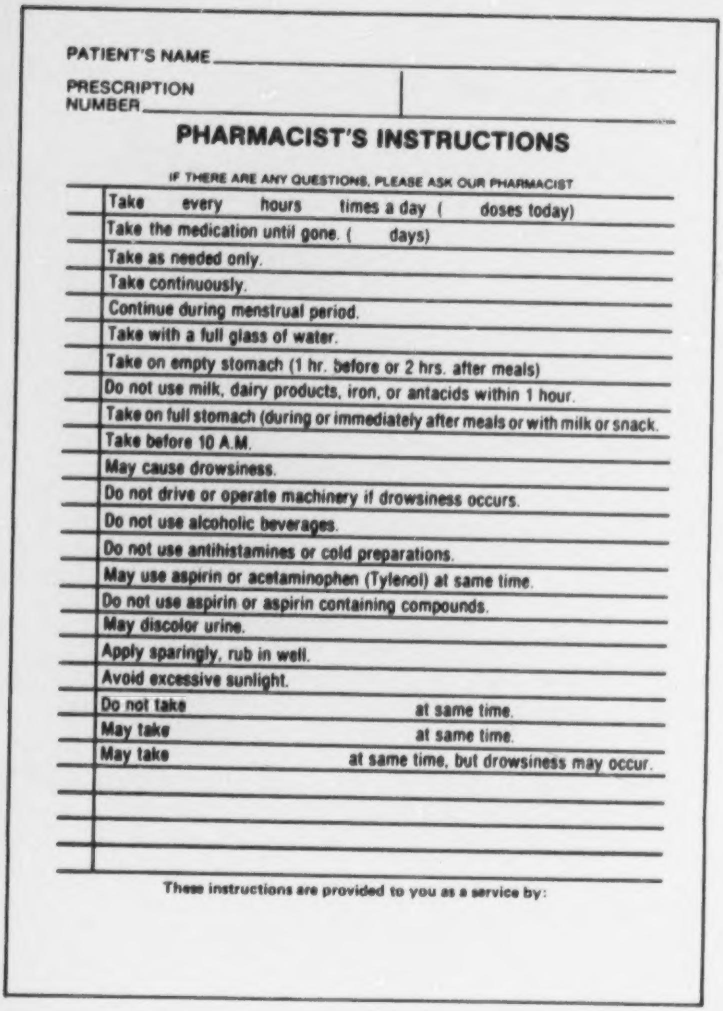
Factors that explain why respondents reported receiving verbal and/or written instructions from the pharmacist inappropriate to the planned educational intervention include: random deviations from the design of the interventions by the pharmacists during the study*; the limitations of the survey instrument; and the dispensing practices ordinarily followed in both study pharmacies. For example, the survey instrument did not specify exactly what was meant by "instructions" received at the time of dispensing. This ambiguity might easily have caused respondents to interpret the typed directions on the prescription container as "written instructions," or the pharmacist's response to questions asked at the time of dispensing as "verbal instructions." Also, as a standard dispensing practice at the study pharmacies, pharmacists routinely counsel patients at a separate, semiprivate section of the transaction counter marked "For Prescriptions Only" and, as needed, provide prescription patients with a printed checklist of special instructions (Fig. 1). Respondents accustomed to such verbal and written instructions could easily have reported receiving them, even though they were not provided during the study. Clearly, however, only respondents in Interventions II and III actually received the study PPIs, and only respondents in Intervention III received the pharmacist-initiated, verbal consultation regarding the PPI and its contents.

Information Desired

As illustrated in Table 2, 63 per cent of the respondents in Intervention I reported wanting more information about their prescriptions than had been received. This group includes 79 per cent of those Intervention I respondents who reported receiving no information of any kind and 42 per cent of those who reported receiving some information. These respondents desired a wide range of information including: the drug's potential for interaction with other drugs and alcohol (31 per cent); specific dosing instructions, such as whether to administer the drug around the clock, during waking hours only or at mealtimes (29 per cent); any special instructions or precautions that should be followed dur-

---

* The incidence of such deviations was probably minimal, since the pharmacists appeared genuinely interested in participating in the study, received training in the parameters of their role only before each intervention, and were given printed reminder sheets for the dispensing counter for each intervention so as to reduce confusion.



PATIENT'S NAME ____

PRESCRIPTION NUMBER ____

PHARMACIST'S INSTRUCTIONS

IF THERE ARE ANY QUESTIONS, PLEASE ASK OUR PHARMACIST

- ____ Take every hours times a day (doses today)
- ____ Take the medication until gone. (days)
- ____ Take as needed only.
- ____ Take continuously.
- ____ Continue during menstrual period.
- ____ Take with a full glass of water.
- ____ Take on empty stomach (1 hr. before or 2 hrs. after meals)
- ____ Do not use milk, dairy products, iron, or antacids within 1 hour.
- ____ Take on full stomach (during or immediately after meals or with milk or snack.
- ____ Take before 10 A.M.
- ____ May cause drowsiness.
- ____ Do not drive or operate machinery if drowsiness occurs.
- ____ Do not use alcoholic beverages.
- ____ Do not use antihistamines or cold preparations.
- ____ May use aspirin or acetaminophen (Tylenol) at same time.
- ____ Do not use aspirin or aspirin containing compounds.
- ____ May discolor urine.
- ____ Apply sparingly, rub in well.
- ____ Avoid excessive sunlight.
- ____ Do not take ____ at same time.
- ____ May take ____ at same time.
- ____ May take ____ at same time, but drowsiness may occur.

These instructions are provided to you as a service by:

FIG. 1. Printed Checklist of Special Instructions. As a standard dispensing practice at the study pharmacies, pharmacists may provide patients with this checklist. During the study this form was not used for Educational Interventions I, II or III.

ing therapy (17 per cent); side effects (17 per cent); and the drug's name, expiration date and indications for use (6 per cent).

In contrast, only 18 per cent of the respondents in Educational Intervention II and 14 per cent of those in Intervention III reported a desire for more information than they received (Table 2). The difference between the proportion of respondents in Intervention I who wanted more information and the proportion in Interventions II and III combined was found to be highly significant ($Z = 6.37, p < 0.01$). This indicates that providing patients with a PPI or a PPI with verbal consultation by the pharmacist significantly reduced their desire for more information about the prescribed drug. The difference between respondents in Interventions II and III who wanted more information is small and could be due to chance.

Of those respondents in Intervention II who reported receiving as well as reading the PPI, 6 per cent wanted additional information, while only 4 per cent of Intervention III respondents who reported receiving and reading the PPI as well as receiving the pharmacist's consultation wanted additional information. These data increase the significance of the difference between Intervention I and Interventions II and III combined regarding the proportion of respondents who desired more information about their prescribed drug ($Z = 7.88$, $p < 0.01$). Once again, the difference in the proportion of respondents desiring more information in Interventions II and III is small and could be due to chance (Table 2).

Respondents who reported experiencing Interventions II and III as planned but who still wanted more information said they wanted to know: what specific infections the drug is used to treat; how best to administer the liquid dosage forms to an uncooperative child; the potential for food–drug interactions; and the manifestations of severe, allergic reactions to the drug.

Knowledge Measures

In addition to satisfaction items regarding the PPI and the pharmacist's verbal role, the interview included the following knowledge items:

1. Should you take this medicine if you are allergic to penicillin?
2. Is it necessary to take all the antibiotic even though you feel better before it is finished?
3. Should this antibiotic be taken with meals?

Respondents who had received the PPI alone (Intervention II) or with verbal counseling by the pharmacist (Intervention III) answered these questions correctly more often than did respondents who had received neither (Intervention I). In addition, although not statistically significant, there was a tendency for knowledge responses to increase in accuracy as the educational intervention increased from I to III.

In response to the first question, 79 per cent of Intervention I respondents answered correctly that the study drugs should not be taken by persons allergic to penicillin (Table 3). This percentage rose to 97 per cent in both Interventions II and III. The difference is significant ($Z = 3.89$, $p < 0.01$) and indicates that providing respondents with a PPI alone or with verbal consultation by the pharmacist significantly increased their correct responses to this item.

The necessity for completing the therapy regimen was known by 85 per cent of the respondents in Intervention I, 87 per cent of respondents in Intervention II and

TABLE 2. Per Cent of Respondents Desiring More Drug Information Than Received by Type of Educational Intervention

Desire More Information	Educational Intervention[*] (Per Cent of Total)		
	I	II	III
Total group	63 (N = 60)[†]	18 (N = 62)	14 (N = 62)
Adjusted group[‡]	63 (N = 60)[†]	6 (N = 49)[§]	4 (N = 48)[‖]

[*] See Table 1 for a description of the Educational Interventions.

[†] Excludes two nonrespondents.

[‡] Includes respondents who reported receiving Educational Interventions II and III as planned.

[§] Excludes 11 respondents who reported not receiving the PPI and two who reported not reading the PPI.

[‖] Excludes five respondents who reported not receiving the PPI, five who reported not reading the PPI and four who reported not receiving the pharmacist's consultation.

TABLE 3. Per Cent of Correct Responses* to Knowledge Items by Type of Educational Intervention

Knowledge Items	Educational Intervention† I	II	III
1. Should take medication if allergic to penicillin?	79	97	97
2. Necessary to take the total prescription?	85	87	92
3. Should take medication with meals?	33	87	89
Total knowledge items	66	90	93

* The denominator for Intervention I excludes one nonrespondent but is 62 for Interventions II and III.
† See Table 1 for a description of the Educational Interventions.

92 per cent of respondents in Intervention III. These differences are small and could be explained as occurring at random in this study.

When asked whether the study drugs should be taken with meals, only 33 per cent of respondents in Intervention I knew that the antibiotics are best administered on an empty stomach (Table 3). By contrast, 87 per cent of the respondents in Intervention II and 89 per cent of the respondents in Intervention III responded to this question correctly. The difference between the proportion of respondents who correctly answered this question in Intervention I and in Interventions II and III combined was found to be highly significant ($Z = 7.57$, $p < 0.01$), which indicates that providing respondents with a PPI alone or with verbal consultation by the pharmacist significantly increased their knowledge regarding this item. Here too, the difference between the proportion of correct responses in Interventions II and III is small and could be explained as occurring by chance.

Attitude Measures

The PPIs enjoyed wide readership and elicited favorable comments by most respondents who received them. Ninety-four per cent of those who reported receiving PPIs also reported reading them. This group includes 96 per cent of the respondents in Intervention II and 91 per cent of the respondents in Intervention III (Table 1). In both interventions, 90 per cent of those who reported receiving and reading the PPI rated its helpfulness positively—64 per cent rating it "very helpful" and 26 per cent "somewhat helpful" in understanding how to take the medication (Table 4). Furthermore, 95 per cent of the respondents who reported receiving and reading the PPI reported satisfaction with the amount of information it contained. All respondents who reported receiving as well as reading the PPI rated its readability positively—90 per cent describing the PPI as "very easy to read" and 10 per cent as "somewhat easy to read." In addition, 99 per cent of the respondents in Educational Interventions II and III who reported that they had received and read the PPI said that they wanted written drug information with other prescriptions. This group includes 98 per cent of the respondents in Intervention II and all the respondents in Intervention III who reported receiving and reading the PPI.

The verbal consultation provided by the pharmacist was also well received by the

TABLE 4. Respondents Reporting Positive Attitudes Regarding PPI by Type of Educational Intervention

	Educational Intervention* (Per Cent of Total)†		
Attitude	II	III	II and III Combined
Helpfulness of PPI	90	91	90
Completeness of PPI	94	96	95
Readability of PPI	100	100	100
Desire PPIs for other prescriptions	98	100	99

*See Table 1 for a description of the Educational Interventions.

† Based on only those respondents who reported reading the PPI: 49 for Intervention II and 53 for Intervention III.

respondents in Intervention III (Table 5). Of those who reported receiving the pharmacist's counseling, 89 per cent found it "very helpful" and 11 per cent "somewhat helpful" in understanding how to take the study drugs. This same group was unanimous in reporting satisfaction with the amount of information received from the pharmacist. Ninety-eight per cent reported that they would like the pharmacist's verbal consultation each time they had a prescription dispensed; 2 per cent would restrict the consultation to new prescriptions only.

Compliance Measures

Behavior in each of the three educational interventions was explored by examining respondents' compliance with the therapy regimen. During the telephone interview, respondents were asked to count the number of doses remaining in their medication containers, as a way of gauging their adherence to the therapy regimen. Since it soon became apparent that respondents who were taking or administering liquid dosage forms could not, by their own admission, reliably estimate the volume of liquid or the number of teaspoonfuls remaining in the bottle, the study of compliance was limited to the 74 per cent of respondents (N = 137) who were taking or administering either tablets or capsules. During the course of the telephone interview, 95 per cent of these 137 respondents agreed to count the number of tablets or capsules remaining in their containers.

Information regarding the beginning date of therapy, the number of days' supply of drug, the daily dosage schedule, and the total number of doses prescribed was available from each respondent's prescription. In each case, the determination of compliance or noncompliance involved comparing the pill count, as reported by the respondent, with the number of doses that should have been in the container if the therapy regimen had been followed correctly up to the date of interview. Non-

TABLE 5. Respondents Reporting Positive Attitudes Regarding the Pharmacist's Verbal Consultation Provided in Intervention III

Attitude	Educational Intervention III* (Per Cent of Total)†
Helpfulness of counseling	100
Completeness of counseling	100
Desire for counseling with other prescriptions	98

* See Table 1 for a description of Educational Intervention III.

† Based on only the 55 respondents in Intervention III who reported receiving counseling; these 55 represent 89 per cent of the 62 respondents who should have received it.

compliance was defined as overutilization or underutilization of the study drugs by the date of the telephone interview.

Overutilization. If the pill count showed fewer doses remaining in the prescription container than there should have been by the end of the interview day, the respondent was classified as noncompliant by virtue of overutilization. For example, a pill count of ten unconsumed doses on the evening of the sixth day of a 30-dose regimen (i.e., a ten-day, three-times-daily therapy) would be classified as overutilization, since no more than 18 doses should have been consumed by the end of the sixth day. This would leave 12 capsules to be consumed during the last four days, instead of ten.†

Underutilization. If the pill count showed that more doses remained in the prescription container than there would have been had the regimen been followed correctly, a different calculation was used to classify the respondent as noncompliant by virtue of underutilization. Not knowing exactly when the therapy should have been initiated, or the time of day when the prescription had been dispensed, the assumption was made that the respondent might still be in compliance with the regimen even if only the last of the first day's required doses was administered. It was further assumed that the last dose of the day could have been scheduled for administration after the time of the telephone interview. For example, a pill count of 15 unconsumed doses on the evening of the seventh day of a ten-day, four-times-daily therapy (40 doses prescribed) could not be classified as underutilization, since the patient might have consumed one dose on the first day of therapy, 20 doses over the next five days and three doses on the interview date up to the time of the interview, for a total of 24 doses consumed. A pill count of up to 16 unconsumed doses at the time of interview could be considered compliant behavior in this case, while a pill count of 17 or more unconsumed doses would require classification as underutilization.‡

Compliance. Using these definitions to determine compliant and noncompliant behavior, analysis shows that 48 per cent of the respondents in Educational Intervention I complied with the therapy regimen (Table 6). The percentage increased to 57 per cent in Intervention II and to 72 per cent in Intervention III.

Although the differences between the percentages of compliant respondents in Interventions I and II, and Interventions II and III, were found not to be statistically significant, there was a tendency for compliance to increase as the rates of educational intervention increased from I to III. The difference between the percentage of compliant respondents in Interventions I and III was found to be significant ($Z = 2.22$, $p < 0.05$) and indicates that compliance increased among respondents who received both a PPI and consultation by the pharmacist, when compared with those who received neither intervention.

Conclusions

The study has demonstrated that patients who lack substantive information about the drugs prescribed for them desire more drug information. Only 37 per cent of the respondents who had prescriptions dispensed with neither a PPI nor planned verbal intervention by the pharmacist were satisfied with those conditions. In the study, respondents preceived the adequacy of the information they received along a kind of continuum, as follows:

---

† Overutilization may thus be stated as $P < N-df$,
where P = the self-reported pill count; N = the number of doses prescribed; d = the day of therapy on which the interview date falls; f = the daily dosing schedule.

‡ Underutilization may be stated as:
$P > N - (df-4)$ for four-times-daily regimens.
$P > N - (df-3)$ for three-times-daily regimens.
$P > N - (df-2)$ for twice-daily regimens;
where P = the self-reported pill count; N = the number of doses prescribed; d = the day of therapy on which the interview date falls; and f = the daily dosing schedule.

TABLE 6. Noncompliant Versus Compliant Behavior With Nonliquid Therapy Regimens by Type of Educational Intervention

Behavior Type	Educational Intervention*		
	I	II	III
Underutilization	13 (33%)	8 (18%)	6 (13%)
Overutilization	8 (20%)	11 (25%)	7 (15%)
Total noncompliance	21 (53%)	19 (43%)	13 (28%)
Total compliant	19 (48%)	25 (57%)	33 (72%)
Total respondents†	40 (101%)	44 (100%)	46 (100%)

* See Table 1 for a description of the Educational Interventions.

† Omits two nonrespondents in I, four in II and one in III who received nonliquid therapy regimens.

1. Among those who reported receiving and reading the PPI as well as receiving the pharmacist's consultation, 96 per cent did not desire additional drug information.
2. Among those who reported receiving and reading the PPI without the pharmacist's consultation, 94 per cent did not desire additional drug information.
3. Among those who reported receiving the PPI and the consultation of the pharmacist (includes those who may not have read the PPI), 86 per cent did not desire additional drug information.
4. Among those who reported receiving the PPI (includes those who may not have read it) without the pharmacist's consultation, 84 per cent did not desire additional drug information.
5. Among those who reported receiving less information in some form, 58 per cent did not desire additional drug information.
6. Among those who reported receiving no information, 21 per cent did not desire additional drug information.

It appears, therefore, that the communication of drug information through the medium of the PPI satisfies patients' desires for information about prescription drugs.

Although some have argued that patients will not read PPIs, this study has demonstrated that patients do read them. Ninety-four per cent of patients aware of the PPI in this study reported reading it. Awareness of the PPI was significantly greater among patients who received it with the pharmacist's consultation (94 per cent aware), though awareness was high (82 per cent) even among patients who received the PPI without reinforcement by the pharmacist.

The study found that knowledge of certain points of information concerning the prescribed drugs was more prevalent among respondents who received the PPI than among those who did not. For two of the three knowledge items, significantly greater numbers of respondents who received PPIs knew the correct answer. It appears, therefore, that PPIs designed with patients' needs in mind can increase their knowledge of prescription drug information.

This study proved to be no exception to the general finding that patients react favorably to PPIs. Ninety per cent or more of respondents who read the PPI reported that they found the information it contained helpful in explaining how to administer the study drugs, complete and readable, and would like to see PPIs dispensed with other prescription drugs. The same kind of enthusiasm has been reported previously in studies of women who received a PPI with oral contraceptives,[8,9] although such women have shown rather disappointing knowledge of the PPI when questioned about its contents.[9,10] It has been suggested that while patients respond favorably to being given PPIs, the reading sophistication required for easy comprehension keeps many from completely understanding the information conveyed and explains why they cannot correctly answer questions about content.[11] The PPIs used in this study, therefore, were designed with care to enhance both usefulness and comprehensibility.

The study also found broad support for counseling by the pharmacist. The planned verbal consultation of Intervention III was well received by respondents; those who reported receiving the pharmacist's counseling were unanimous in rating the information conveyed as complete and helpful. Only one respondent did not want the pharmacist's advice given routinely with prescription drugs, and that person suggested having the pharmacist counsel only patients with new prescriptions and not those with refills. Moreover, the pharmacist's counseling was instrumental in increasing awareness of the PPI among patients.

The findings of the study also support the conclusion that PPIs can enhance patient compliance with prescribed antibiotic regimens. The results show that patients given PPIs and verbal reinforcement by the pharmacist have significantly greater compliance rates than patients who receive neither.

Acknowledgments

Appreciation is extended to the Hunterdon Health Fund for the financial support of this project. In addition, the authors gratefully acknowledge the technical expertise provided by the following individuals: Nancy H. Bryant, R.N., M.P.H., Maria Cutie, M.S., James Beatty, Pharmacist, Edward Malone, Pharmacist, Myra Madnick, M.Ed., Rose Donaldson, M.S., Maxine Susman, Ph.D., Phyllis Supino, Ed.D., Teri Manes, M.P.H., Carol Wimmer and Terri Bilheimer.

References

1. Morris LA, Halperin JA. Effects of written drug information on patient knowledge and compliance: a literature review. Am J Public Health 1979;69:47.

2. Green L, Faden RR. The patient package insert: potential effects on the patient—3. Drug Inf J 1977;11:64S.

3. Stewart RB, Cluff LE. A review of medication errors and compliance in ambulant patients. Clin Pharmacol Ther 1972;13:463.

4. Davis MS. Physiologic, psychological and demographic factors in patient compliance with doctor's orders. Med Care 1968;6:115.

5. Becker MH, Green LW. A family approach to compliance with medical treatment: a selective review of the literature. Int J Health Educ 1975;18:173.

6. Dwyer RF, Hammel R. An experimental study: patient package inserts and their effects on hypertensive patients. Urban Health 1978;7:46.

7. Clinite JC, Kabat HF. Improving patient compliance. J Am Pharm Assoc 1976;16:74.

8. Weibert RT. The patient package insert: potential distribution problems—1. Drug Inf J 1977;11:45S.

9. Fleckenstein L, Joubert P, Lawrence R, et al. Oral contraceptive patient information: a questionnaire study of attitudes, knowledge and preferred information sources. JAMA 1976;235:1331.

10. Ryan JR, McMahon FG. The patient package insert: potential effects on the health care system—2. Drug Inf J 1977;11:75S.

11. Liguori S. A quantitative assessment of the readability of PPIs. Drug Intell Clin Pharm 1978;12:712.

Appendix

Penicillin

Your penicillin may be known by one of several brand names such as Bicillin®, Pentids®, Pen Vee K®, V-Cillin® or others.

Pencillin is an antibiotic used to treat infections caused by certain bacteria. As a rule, PERSONS WHO HAVE HAD AN ALLERGIC REACTION TO PENICILLIN SHOULD NOT TAKE IT AGAIN.

- If you are allergic to penicillin and have not told your doctor, tell him before starting penicillin. If you have a history of allergies to many things and have not told your doctor, tell him/her before starting penicillin.

How to Take Penicillin by Mouth

- Take *all* your penicillin. If the doctor orders a refill, you should *finish it all* too. Don't stop taking penicillin after a few days because you feel better. Continue taking it because the infection may still be in your body.
- Follow the instructions on the label and take every scheduled dose of pencillin with a glassful of water. DON'T SKIP ANY DOSES. If you have a question about the number of doses to take or the timing of each dose, ask your pharmacist.
- *Whenever possible* take each dose of pencillin on an empty stomach ONE HOUR BEFORE OR TWO HOURS AFTER MEALS. This is because food, fruit juices, or antacids like Gelusil®, Maalox®, Alka Seltzer®, etc. may slow down the effect of penicillin. When it's not possible or convenient to do this, take each dose of penicillin anyway. Remember, DON'T SKIP ANY DOSES.

Advice to Patients

- If you begin to have any problems while taking penicillin, such as itching, skin rash or hives, nausea, vomiting or diarrhea, call your doctor.
- Do not give any of this medicine to anyone else. Keep this and all medicines out of the reach of children.
- Penicillin tablets should be stored at room temperature away from excess moisture.
- Penicillin liquid, for teaspoonful or dropper dosage, should be stored in a refrigerator. Shake well before taking. After 14 days amounts not used should be thrown away (flushed).
- If you have any questions about penicillin, ask your doctor or pharmacist.

Ampicillin

Your ampicillin may be known by one of several brand names such as Polycillin®, Principen®, Penbritin® or others.

Ampicillin is an antibiotic used to treat infections caused by certain bacteria. Ampicillin is like penicillin. So, as a rule, PERSONS ALLERGIC TO PENICILLIN SHOULD NOT TAKE AMPICILLIN.

- If you are allergic to penicillin or ampicillin and have not told your doctor, tell him/her before starting ampicillin.

- If you are pregnant, or think you might be, and have not told your doctor, tell him/her before starting ampicillin.

How to Take Ampicillin by Mouth

- Take *all* your ampicillin. If the doctor orders a refill, you should *finish it all* too. Don't stop taking ampicillin after a few days because you feel better. Continue taking it because the infection may still be in your body.
- Follow the instructions on the label and take every scheduled dose of ampicillin with a glassful of water. DON'T SKIP ANY DOSES. If you have a question about the number of doses to take or the timing of each dose, ask your pharmacist.
- *Whenever possible* take each dose of ampicillin on an empty stomach ONE HOUR BEFORE OR TWO HOURS AFTER MEALS. This is because food, fruit juices, or antacids like Gelusil®, Maalox®, Alka Seltzer®, etc. may slow down the effect of ampicillin. When it's not possible or convenient to do this, take each dose of ampicillin anyway. Remember, DON'T SKIP ANY DOSES.

Advice to Patients

- If you begin to have any problems while taking ampicillin such as itching, skin rash or hives, nausea, vomiting or diarrhea, call your doctor.
- Do not give any of this medicine to anyone else. Keep this and all medicines out of the reach of children.
- Ampicillin capsules and tablets should be stored at room temperature away from excess moisture.
- Ampicillin liquid, for teaspoonful or dropper dosage, should be stored in a refrigerator. Shake well before taking. After 14 days amounts not used should be thrown away (flushed).
- If you have any questions about ampicillin, ask your doctor or pharmacist.

MEDICAL CARE
June 1982, Vol. XX, No. 6

A Survey of Patients' Receipt of Prescription Drug Information

LOUIS A. MORRIS, PH.D.

A nationwide telephone survey of 1,223 individuals investigated the nature and source of information provided to patients regarding prescription drugs. About half of the respondents said that they had received information from their doctor about the purpose and directions for use for their most recent prescription. Only a few people (11 per cent) said that they had been informed about the drug's side effects, and 19 per cent said that they had been told nothing by their doctor. Most (72 per cent) related that nothing had been said to them at the pharmacy. Written information (stickers on the medicine container, leaflets or brochures) was said to be infrequently provided at the pharmacy. About 12 per cent of the respondents said they had expected they might get a drug side effect, but only 9 per cent said that they had experienced one. The most frequently cited action in response to side effects was to consult the physician (40 per cent); however, a sizable percentage of people stopped the medicine completely or temporarily (36 per cent) or kept on taking the drug as prescribed (32 per cent).

THE MOST COMMON FORM of medical treatment rendered by physicians is the prescribing or administering of a drug. Two thirds of the ambulatory therapeutic services given in physicians' offices involve the use of prescription drugs.[1] According to a recent survey, about a quarter of the adult U.S. population took a prescription drug during the past 24 hours.[2]

Numerous scientific journals, academic and government publications, and drug industry promotional materials provide a constant flow of information to prescribers and dispensers of these medicines. In this manner, professional decisions can be made on the basis of the best scientific information available. However, the flow of information to consumers about prescribed medicines is considerably more confined. By legal definition, a drug is marketed on a prescription basis, rather than made available over the counter, because of its potential to cause harm.[3] To avoid dangerous interactions with foods or other drugs and to avoid hazardous activities, it may be argued that consumers need to know certain precautions about prescribed medicines. Also, prescription medicines often have side effects associated with their use. Providing consumers with information about side effects may allow better recognition, interpreta-

---

From the Food and Drug Administration, Rockville, Maryland.

The data collection for this study was funded by FDA contract 223-79-8250.

The views expressed are solely those of the author and do not reflect the policy of the Food and Drug Administration.

Address for reprints: HFD-175, Food and Drug Administration, 5600 Fishers Lane, Rockville MD 20857.

0025-7079/82/0505/0596/$01.00 © J. B. Lippincott Co.

tion and knowledge about preventive or remedial action.

In addition to precautionary and side-effect information, consumers also need proper instruction on how to use drugs so that medicines will produce their intended benefits. Several literature reviews suggest that 30 to 50 per cent of patients do not adhere to prescribed regimens,[4–6] frequently missing doses, consuming too much or too little medicine, prematurely discontinuing the regimen or improperly spacing doses. Although there are many reasons for this large-scale failure to adhere to the regimen, a significant proportion of the problem can be traced to inadequate communication between physicians and patients.[7,8]

For individuals using prescription medicines on an outpatient basis, the physician is often viewed as the primary expected source of information. Studies done in this country and abroad reveal a wide range of estimates of the amount and type of information provided to patients.[7–13] Much of this variance may be attributable to the nature of the population under study (i.e., age, disease type, patient status, treatment facility, etc.) and the method utilized to collect data.

For example, Svarstadt,[7] using direct observational techniques, found that physicians in a New York clinic informed patients about the length of therapy 10 per cent of the time and dosage frequency 17 per cent of the time. For 17 per cent of the patients, the drug was never discussed. In a telephone survey, Hoff[9] found that 48 per cent of 415 respondents reported that they had not talked to their doctor when their most recent prescription was written. Boreham and Gibson interviewed 80 Australian female patients.[10] During initial consultation, physicians provided information on a drug's name, purpose and effects, and gave instructions for treatment to about one fifth to one third of the patients. Information about side effects was provided to 15 per cent of the patients; however, this disclosure was most frequently spurred by patient questioning. During follow-up consultations, about half of the patients were apt to be informed about a drug's name, but other information was infrequently provided. Taken as a whole, these studies suggest far less than optimal communication of prescription drug information from physician to patient.

Another potential source of information for the patient is the pharmacist. Pharmacists are in a central position in the drug distribution system to inform patients about prescription drugs, and numerous studies have demonstrated clear benefits of pharmacist counseling. However, studies that have used direct observational methods reveal that only 0 to 20 per cent of patients filling prescriptions receive meaningful information about their prescription.[14–16] A study of 95 midwestern pharmacies[17] found that in 41 per cent of the pharmacies some oral or written information was provided when the drug was dispensed, but the quality of information was rated as poor. Campbell and Grisafie[18] found that in Washington State, where regulations require pharmacists to counsel patients, about half of the pharmacists studied provided information on how to use a specific product, but only 8 per cent provided an important precaution about use of the drug. In addition to oral sources of information, there are a number of printed sources potentially available to the consumer.[19] These sources range from small stickers to be affixed to the medication container to multipage borchures and pamphlets.

Previous studies provide a glimpse of the information environment surrounding consumers' use of prescription drugs. Due to regional differences, small samples and vast differences in data collection methods, the emerging picture (especially that part involving physician delivery of drug information) is sketchy.

The purpose of the present study was to provide a more complete overview of consumers' receipt of prescription drug information. How much and what type of information is being provided by health professionals? What other sources of information (friends, relatives, printed media) provide information to consumers? A second purpose was to provide data from a broad cross section of individuals regarding how information influences drug use experience, specifically the experience of drug side effects. What proportion of the people taking drugs experience a side effect? Is the side effect expected, and what is done in response to its occurrence?

To gather information in a cost-feasible manner, a national telephone survey was undertaken. This method enables the solicitation of a wide range of information from a broad cross section of the population. Several studies have found that telephone interviews yield generally comparable data to personal interview and mail surveys.[20-22] Many suspected forms of bias related to telephone interviews, such as restricted access to telephones by certain population segments and unwillingness to answer sensitive questions over the phone, have not been confirmed in systematic evaluations. However, other forms of response artifacts have been shown to influence responding. For example, the pacing of questions and the time allowed for recall of events may influence subject reporting. For example, Siemiatycki[20] found less underreporting of doctor visits when respondents answered questions via mail survey compared with telephone or personal interviews. Thus, in the present survey care was taken to fix an experience in the subject's mind (the last time a prescription was filled) in the beginning of the survey and to ask a series of questions related to that experience. This form of questioning has the advantage of soliciting information on reported experience, rather than global attitudes or opinions. However, its major disadvantage is that it relies solely on the subject's memory. Respondents may have forgotten, misperceived or incorrectly recalled events or details of events. Thus, results must be carefully interpreted as based on subject memory, and study conclusions must be couched as tentative, to be confirmed by observational studies.

Method

A national (continental U.S.) survey of telephone households was conducted in the summer of 1979. Selection of telephone numbers was accomplished via a three-stage sampling technique. At the first stage, each of the 30,000 telephone exchanges (the first three digits of the telephone number) currently in use was assigned to one of nine census tract sampling strata. Exchanges were then ranked by the size of the population covered and placed in sampling blocks of 55 exchanges. Two telephone exchanges were then chosen at random from each block.

In the second sampling stage, telephone households were identified via a computer-generated random four-digit suffix. Prior to issuing the selected telephone suffix, a computer program checked the selected number against information in storage regarding whether the selected number was a residential telephone number. The fourth and fifth digit of the telephone number are coded separately for businesses and residential phones. In this manner, approximately 70 per cent of the nonworking or business numbers were eliminated before the telephone number was identified for the interviewer.

The third stage of the process consisted of selecting the eligible survey respondent within the household. Individuals 18 years of age and older were considered eligible for the study. Interviewers obtained information from the individual answering the telephone on the number of adults (over the age of 18) in each contacted house. By virtue of a precoded random schedule, one individual was selected for interviewing in each household.

To improve response rates, up to five calls, distributed on day, nights and weekends, were made to each selected number. People who were too busy to be interviewed were called a second time and asked to complete the interview. In addition, a nonresponse follow-up was conducted on a sample of numbers for which there was no answer, busy numbers and numbers for which the selected individual was not at home. The follow-up consisted of a second round of four telephone calls. The noncontacted telephone numbers were reviewed with the telephone company to classify them as working, nonworking or nonhousehold. Table 1 outlines the sample dispositions. Of 3,580 telephone numbers called, 1,610 households were identified. In all, 1,223 subjects completed the survey; 1,180 were questioned on the initial round of telephone calls, and 43 subjects were added by the nonresponse follow-up. This yields an overall response rate of 76 per cent.

Interviewers for the study had experience with previous telephone surveys. All were familiar with interviewing techniques, and all were briefed on the general purpose and on specific questions for this study. Role-playing techniques were used to identify possible problems prior to the study, and interviewers were periodically monitored by interviewing supervisors throughout the data collection period.

TABLE 1. Sample Selection Dispositions

Telephone numbers called	3,580
Nonhousehold (business, nonworking)	1,572
No answer or busy	398
Identified households	1,610
Ill, indisposed	12
Language barrier	75
Refusals	236
Call backs (for identified respondents)	78
Original respondents	1,180
Those added by follow-up	43
Total respondents	1,223

Questionnaire Design

Those individuals consenting to be interviewed were told that the purpose of the study was for the Food and Drug Administration to find out more about their experiences with medicines. To place the respondent in an experiential rather than an attitudinal frame of reference, respondents were first asked to recall the last time they had a prescription filled for themselves. Interviewers then proceeded to ask a series of questions about the respondents' experiences filling the prescription; what respondents were told about the medicine(s) from their doctor, pharmacist and friends; and whether they received any written information. Data on experience using the drug and reasons for taking the drug were solicited, and a question about educational background was posed. A list of the questions posed is presented in the Appendix. Subjects who stated that they never had a prescription filled for themselves (N = 48) and those who did not remember (N = 62) were asked to talk about their experiences during the last year filling a prescription for another family member. Fifty-four respndents remembered filling a prescription for a family member. Since about half (N = 21) of these prescriptions were for a child, these data were considered along with those for respndents who remembered filling prescriptions for themselves.

Results

Subjects in the sample ranged in age from 18 to 92 years, with a mean age of 44. Fifty-eight per cent were female. About 78 per cent had graduated from high school, and 41 per cent had at least some college education.

Subjects were asked a series of questions describing the circumstances when their most recent prescription was filled. Three fourths said that they had filled a prescription for a single medication at that time, and most others (15 per cent) said that at least some of the drugs had been taken on a previous occasion, with 41 per cent of these people refilling an existing prescription. Most of the respondents (83 per cent) said that they themselves had picked up the medicine at the pharmacy, and 62 per cent of the time the medicine was said to have been given to them by a pharmacist rather than by a clerk. Approximately 12 per cent said that they had filled their most recent prescription within the preceding week, 18 per cent one to four weeks previously, 30 per cent one to six months previously, 17 per cent six to twelve months previously and 25 per cent more than a year previously.

Overview of Information Provided

Subjects were asked to recall freely what, if anything, they had been told by the person who sold them the medicine(s). Up to four responses were recorded for each respondent. Table 2 lists the type of information provided. Seventy-two per cent of the people answering this question said that they had not been told anything about the medicine. The most frequently provided information concerned directions about how and when the medicines should be used. Side-effect information was infrequently provided. However, one in five people taking headache remedies or birth control pills said that they had been told about side effects at the pharmacy.

In addition to verbally supplied information, 8 per cent (73 respondents) said they had been given written material in addition to the label on the medicine container. About half (41 per cent) of the provided information was said to be stickers for the medicine bottle, with the rest receiving brochures, pamphlets or instruction sheets. About half (55 per cent) of the 47 people reporting that they had received longer forms of written information said they were taking drugs for menopause or birth control, cases for which written information is required to be dispensed by federal regulation.

To ascertain physician-supplied information, patients were asked to relate what, if anything, their doctor told them about the drug(s) when it was prescribed or while they were taking it. Again, up to four re-

TABLE 2. Oral Information Provided at Pharmacy

Information	Number	Per Cent*
Directions for use	196	21
Side effects	42	5
Purpose of medicine	44	5
General (expiration date, refills, price)	48	5
Nothing was said	673	72
Total	1,103	

* Sums to over 100%; 932 people answered this question, up to four responses were recorded for each.

TABLE 3. Oral Information Supplied by Physician

Information	Number	Per Cent*
Purpose	561	51
Directions—drug use	516	47
Directions—call/ contact physician	56	5
Side effects	123	11
General (refills, read pamphlet)	27	2
Nothing was said	207	19
Total	1.490	

* Sums to over 100%; 1,101 people answered this question, up to four responses were recorded for each.

sponses were recorded. Table 3 lists the results for this question.

Clearly, subjects related that they were told a great deal more by their physician than by the pharmacist. About half of the respondents said that they were given information about the purpose of the medicine, and about half said that they were given directions about how to use it. Directions for use included information on when to take doses and how long to continue on the medicine. Only 5 per cent of the 516 subjects who said they received directions for use information related that they were given any precautions about taking the medicine (e.g., take with milk, avoid alcohol, etc.). About one fifth of these subjects said that they were told nothing about the medicine, but a small percentage of them (11 per cent) volunteered that no information was necessary because they were experienced in using the drug.

About one in ten were told about side effects. In 12 per cent of these 123 cases, the information relayed was that there would be no side effects. Side-effect information was most frequently provided to people taking birth control pills (said to be relayed to 46 per cent of those taking this drug). The other drugs for which side-effect counseling was said to be provided to more than about one in five people were drugs used for dieting, menopause, headache and allergies.

Only about 5 per cent of the respondents related that they were told when to contact the physician again. Of those informed, the reason for the contact was fairly evenly divided among three alternatives: a prearranged date (32 per cent); if there were side effects (30 per cent); and if there was no improvement (30 per cent). Only 12 per cent said that they had been given general information, such as refill instructions. This general information category included respondents who said that their physician told them to read the pamphlet that comes with the medicine. This occurred primarily for birth control pills and menopausal drugs; eight of 53 respondents taking these drugs said that their doctor had told them to read the patient package insert.

In addition to the physician and pharmacist, patients may receive information about the drugs they are taking from their friends and relatives. When asked if their friends and relatives told them anything about the medicine, the most frequent answer was "no" by 89 per cent of the respondents. However, 42 per cent of those taking birth control pills and 26 per cent of those taking menopausal drugs said that their relatives did talk to them about the medicine.

Experience of Side Effects

One treatment outcome that may alter the course of drug therapy is the patient's experience of drug side effects. Only 9 per cent of the respondents said that they experienced a side effect when they took their medicine. Twelve per cent of the respondents signified that they had expected a drug side effect before they took the medicine. Forty-six of the 108 people (43 per cent) who had side effects said they had expected one, whereas only 10 per cent of those who did not have a side effect said they had expected one. The expectation of side effects was greatest for those taking medicines for dieting, birth control, headache, allergies and menopause. The experience of side effects was most frequently reported by those taking drugs for dieting, allergies and headache.

When the people were asked how they knew they might get a side effect, the most frequently provided reasons were that they were so informed by their physician or that they had previous experience with the medicine. The only exception was for women taking birth control pills, two thirds of whom said they had read about side effects in a pamphlet. Other reasons, given less frequently, are listed in Table 4.

The respondents who said that they had experienced a side effect were asked what they did about the reaction. Table 5 lists the response to this question. The most frequent reaction was to call the physician, although sizable proportions of the respondents said that they either did nothing or stopped taking the medicine completely. The type of action taken in response to drug side effects differed according to the type of condition being treated. The most common response for those taking a drug to cure an infection was either to stop taking the medicine or to continue on the regimen without consulting their physician. Only one quarter of the 44 people experiencing a side effect called their physician. For those taking medicine for a chronic condition, half of the 30 people experiencing a side effect called their doctor.

TABLE 4. Reasons for Expecting Side Effects

Reason	Number	Per Cent*
Previous experience	41	33
Read about it (newspaper, magazine)	5	4
Read about it (brochure/pamphlet/book)	15	12
Information on bottle	6	5
Doctor	52	42
Pharmacist	8	7
Relative/friend	14	11
Other	6	5
Total	147	

* Sums to over 100%; 123 people answered this question, up to three responses were recorded for each.

Discussion

As mentioned previously, one must be especially cognizant of the role of respondent memory when interpreting the results of this survey. First, respondents may have simply forgotten the last time they filled a prescription. Since only 9 per cent said that they had never filled a prescription or did not know when they filled a prescription, the great majority of respondents recalled some experience, even if it was not their most recent one.

A second, but perhaps more critical, issue is that memory is imperfect. It is likely that subject remembrances are distorted in at least two ways: 1) "leveling," or simply forgetting events or details of events that did indeed occur and 2) "sharpening," magnifying or creating details of events that never occurred. It is impossible to know the extent to which "leveling" and "sharpening" occurred in the present survey. One might expect that tabulating results based on the recency of prescription filling may have helped highlight memory problems. This was attempted, but unfortunately these data are confounded with other factors (such as purpose of the medicine and previous use of the drug) that appear better able to explain the data patterns. Studies have shown that even when interviewed a few minutes after leaviang the physician's consulting room, patients fail to report a third to half of the pertinent information provided at the visit.[23,24] One may presume that respondent memory has an important bearing on the data results,

TABLE 5. Response to Side Effects

Response	Number	Per Cent*
Stopped medicine completely	29	27
Stopped temporarily	10	9
Called doctor	43	40
Called pharmacist	1	1
Kept taking medicine	35	32
Reduced dose	5	5
Rested/relaxed	8	7
Other	2	2
Total	133	

* Sums to over 100%; 108 people answered this question, up to three responses were recorded for each.

regardless of the time lapse between prescription filling and interview. A basic caution that must be applied in interpreting study results is that this survey represents the perception and memory of events rather than direct measurement of events.

With this caution in mind, it is apparent that verbal communication by the physician is the primary method of informing patients about drugs. The drug's purpose and usage instruction are reported to be received by about half of the subjects. It is clear that patients see their doctor as the most important source of drug information.[25,26]

The pharmacy is often considered an ideal place to receive drug information. Most said that the pharmacist personally gave them the medicine, but the majority of the respondents related that they were told nothing about the drug. When advice was given at the pharmacy, it was primarily on how to use the drug. It should be noted that patients were asked to recall physician-provided information given at any time during the course of treatment, whereas pharmacy-provided information was solicited only for the time the prescription was filled. Since only one respondent reported seeking advice from a pharmacist when side effects occurred, it is evident that pharmacist-provided information is most apt to be provided at the time the drug is dispensed. Respondents did report a greater likelihood of being counseled at the pharmacy if they were receiving a prescription they had not taken before. However, these differences were not large, and overall the majority reported receiving no advice. The general picture that emerges is that people remember being told very little about their drugs, with the exception (for half the respondents) of drug purpose and how to take the medicine.

The expectation that one may get a side effect was most likely to be conveyed by the physician or by previous experience using the medicine. The only exception was birth control pills, where two thirds said that information was conveyed in a pamphlet. Interestingly, birth control pills and menopausal drugs were among those drugs for which subjects most frequently reported to have received side-effect information from their doctors. These drugs were also reported as being most often discussed with friends and relatives. Birth control and menopausal drugs were chosen as candidates for patient package inserts (PPIs) because of their risks and the elective nature of their use. The precise role of the PPI in promoting or reducing the amount of discussion about their use cannot be determined from this survey.

Due to the undetermined role respondent memory may have played in influencing the obtained data, this survey's results should be interpreted in light of other studies that used alternate methodologies. The results appear consistent with literature previously reviewed. In addition, a recent survey of pharmacies in 20 U.S. cities found that 72 per cent of the people filling prescriptions for conjugated estrogens did not receive any meaningful verbal counseling from the pharmacist.[27] This is the same percentage as that found in the present survey. Because conjugated estrogens are drugs for which pharmacists may not be as likely to counsel patients as other drugs, this concordance of data must be carefully interpreted in light of the methodological limitation of the present survey. A useful purpose for this data is to provide a baseline for future telephone surveys so that trends in drug education and patient counseling may be discerned.

Acknowledgment

The assistance of Drs. Philip Ley, Patricia McCormick and Evelyn Gordon, and the Chilton Research Service in the design of this study and the preparation of this report are gratefully acknowledged.

References

1. Koch H, Dennison N. Ambulatory medical care rendered in physician offices: United States, 1975. Advancedata 12: HRA 77-1250, 1977.

2. Roper Reports. Section Q, Question 66. New York: The Roper Organization, 1980;80.

3. Food and Drug Administration. Food, Drug and Cosmetic Act, as amended, October 1976. Section 502(f). Washington, D.C.: Government Printing Office, 1976;44.

4. Marston M. Compliance with medical regimens: a review of the literature. Nurs Res 1970;19:312.

5. Barofsky I. Barriers to effective care in the chronic psychiatric patient. In: Barofsky I, Budson R, eds. Chronic psychiatric patient in the community: principles of treatment. New York: Spectrum Press, in press.

6. Sackett D. The magnitude of compliance and noncompliance. In: Sackett D, Haynes B, eds. Compliance with therapeutic regimens. Baltimore: Johns Hopkins Press, 1976:9.

7. Svarstad B. Physician–patient communication and patient conformity with medical advice. In: Mechanic D, ed. The growth of bureaucratic medicine. New York: John Wiley and Sons, 1976:221.

8. Ley P. Psychological studies of doctor–patient communication. In: Rachman D, ed. Contributions to medical psychology. New York: Pergamon Press, 1977:9.

9. Hoff L. How often do consumers seek your advice on Rx and OTC products? Pharm Times 1975;163:52.

10. Boreham P, Gibson D. The information process in private medical consultations: a preliminary investigation. Soc Sci Med 1978;12:409.

11. Fletcher S, Fletcher R, Thomas D, Hamann C. Patients' understanding of prescribed drugs. J Community Health 1979;4:183.

12. Bain D. Patient knowledge and the content of the consultation in general practice. Med Educ 1977;11:347.

13. Hulka B, Cassel J, Kupper L, Burdette J. Communication, compliance, and concordance between physicians and patients with prescribed medicines. Am J Public Health 1976;66:847.

14. Knapp D, Wolf H, Knapp D, Rudy T. The pharmacist as a drug advisor. J Am Pharm Assoc 1969;9:502.

15. Rowles B, Keller S, Gavin P. The pharmacist as compounder and consultant. Drug Intelligence and Clinical Pharmacy 1974;8:242.

16. Cutting corners at the corner drug store: an investigation of pharmacy practices in greater Boston. Report of the Commonwealth of Massachusetts Consumers' Council, Boston, Mass., 1976.

17. Rickett F, White S, Mossberg H, Matchett J. Pharmacist/patient counseling practices. Cont Pharm Pract 1978;1:67.

18. Campbell R, Grisafe J. Compliance with Washington State patient information regulation. J Am Pharm Assoc 1975;15:494.

19. Morris L. A sampler of patient oriented prescription drug materials. Drug Intelligence and Clinical Pharmacy 1978;12:161.

20. Klecka W, Tuchfarber A. Random digit dialing: a comparison to personal surveys. Public Opinion Quarterly 1978;42:105.

21. Lucas W, Adams W. An assessment of telephone survey methodology. Santa Monica, Calif.: Rand Corporation, R-2135-NSF, 1977.

22. Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys. Am J Public Health 1979;69:238.

23. Ley P, Spelman M. Communications in an out-patient setting. Br J Soc Clin Psychol 1965;4:114.

24. Joyce C, Caple G, Mason M, et al. Quantitative study of doctor–patient communications. Q J Med 1968;150:183.

25. Joubert P, Lasagna L. Patient package inserts: II. Toward a rational patient package insert. Clin Pharmacol Ther 1975;18:663.

26. Udkow G, Lasagna L, Weintraub M, Tamoshunas Z. The safety and efficacy of the estrogen patient package insert. JAMA 1979;242:536.

27. Morris L, Myers A, Gibbs P, Lao C. Estrogen PPIs—an FDA survey. Am Pharm 1980;20:22.

Appendix

Questions Asked*

1. How long has it been since the *last time* you had a prescription filled *for yourself*?
2. At that time, when you last had a prescription filled for yourself, was it for only *one* medicine or was it for *more than one*?
3. How many different medicines did you get filled for yourself at that time?

---

* Skip patterns and interviewer instructions are not reported. A second version of the questionnaire was used for respondents filling prescriptions for others. Sex and age of respondents were recorded prior to asking these questions.

4. When you had this (these) prescription(s) filled, was it the *first* time you had ever taken this (these) medicine(s), or had you taken it (them) before?
5. Thinking of the medicine(s) that you had taken before, was (were) the prescription(s) just being refilled by the pharmacy at your request, or did you or the pharmacy have to contact the doctor before having it (them) refilled?
6. Did you pick up the prescription(s) at the pharmacy yourself?
7. Who actually gave you the medicine(s) . . . was it the pharmacist, a clerk, or someone else?
8. What, if anything, did the person who sold you the medicine(s) tell you about the medicine(s)? Did the person say anything else?
9. In addition to the label on the medicine container, were you given any other written information at the pharmacy?
10. What kind of written information was this? Was it . . .
    Brochures or pamphlets
    Instruction sheets
    Stickers on the medicine with instruction, such as "May cause drowsiness, etc."
    Or some other type of written information?
11. What, if anything, did your doctor tell you about the medicine(s) when it was (they were) prescribed for you or while you were taking it (them)? Were you told anything else?
12. Did any of your friends or relatives tell you anything about the medicine(s) that was (were) prescribed for you?
13. Did you have any side effects or reactions when you took the medicine(s)?
14. Did you expect that you might have a side effect or reaction *before* you took the medicine(s)?
15. How did you know that you might have a side effect or reaction?
16. What did you do about the reaction you were having to the medicine(s)?
17. Please tell me what condition the medicine(s) was (were) prescribed for.
18. (A question asking subjects to rank importance of three side effects was included—data are not reported.)
19. And, finally, just one question about yourself. Would you please tell me the last grade of school you completed?

MEDICAL CARE
June 1982, Vol. XX, No. 6

Satisfaction With Ambulatory Care and Compliance in Older Patients

MARGARET W. LINN, PH.D.,* BERNARD S. LINN, M.D.,†
AND SHAYNA R. STEIN, PH.D.‡

Predictors of satisfaction with ambulatory care and compliance in 267 older and 581 younger patients were determined. Each patient rated a 45-item satisfaction-with-care-scale. Race, SES, marital status, distance from clinic, severity of illness (as measured by physician ratings, self-health assessment, number of medications, number of diagnoses, and number of clinic visits and hospitalizations in the prior year), and physician expectations of improvement were entered as predictors into stepwise multiple regression analyses for the elderly and the young. Predictors of better satisfaction in the young were less severe conditions, being nearer to the clinic and having fewer prior clinic visits over the year. In the elderly, having fewer visits to the clinic, more expectation of improvement by the physician and less severe conditions were associated with better satisfaction. Severity and clinic visits were predictors in each age group. The young, however, were also influenced by distance from the clinic. The elderly were influenced separately by the physician's prognosis. Thus, when the more impaired elderly are seen frequently without expecting a benefit, their satisfaction with care is poor. Further, satisfaction with care was correlated significantly with compliance in the elderly but not in the young. Findings suggest that improving satisfaction with care might also improve rates of compliance with the medical regimen in older patients.

SATISFACTION WITH MEDICAL CARE and compliance with a medical regimen are two variables often studied in health care. Generally, studies deal with one or the other variable alone; less frequently, they deal with the association between these two aspects of care. It has been pointed out that satisfaction with care has been viewed both as a determinant and a consequence of utilization.[1] Cafferate,[2] however, noted that the two views are not incompatible, and that feedback and reciprocal relationships may be present. Satisfaction with care has often been included among patient outcomes regarding quality of care,[3,4] and this would appear justified. Compliance has sometimes also been seen as an outcome of care, although it might better be classified as an intermediate variable influencing outcome as measured by disability, death or disease, since compliance and noncompliance may result in differential outcomes along these lines.[5] In some instances, it seems logical that satisfaction with care received would be predictive of whether or not patients complied with the regimens prescribed. In this respect, satisfaction would be an independent, and compliance a dependent, variable.

Some determinants of satisfaction with care have been reported. A trend toward

---

* Director, Social Science Research, Veterans Administration Medical Center, Professor of Psychiatry, University of Miami School of Medicine, Miami, Florida.

† Associate Chief of Staff for Education, Veterans Administration Medical Center, Professor of Surgery, University of Miami School of Medicine, Miami, Florida.

‡ Research Social Scientist, Veterans Administration Medical Center, Miami, Florida.

Address for reprints: Margaret W. Linn, Ph.D., (151), VA Medical Center, 1201 N.W. 16 Street, Miami, FL 33125.

0025-7079/82/0505/0606/$00.95 © J. B. Lippincott Co.

greater satisfaction with care in older persons was found by Gerst,[6] particularly among older males; however, Hulka et al.[7] identified lower satisfaction among patients over the age of 60, particularly in older blacks. Others[8] have reported no relationship between satisfaction with care and demographic variables. Greater severity in regard to illness and disability has been found to accompany lower satisfaction with care,[9,10] and in at least one study,[11] satisfaction was associated with compliance.

Becker et al.[12] concluded that patient noncompliance has become the best documented but least understood health-related variable. Most studies[13,14] have found that approximately a third of their patients failed to follow physicians' recommendations. Adherence to instruction has been reported to be extremely low in low-income patients, for whom noncompliance reached 60 per cent or higher.[15] Age extremes have been found to relate to noncompliance, with the very young resisting bad-tasting medicines and the very old simply forgetting to take their medications.[16] Kasl and Cobb[17] believed that compliance studies had focused too much on demographic variables rather than on attitudes and perceptions in explaining whether or not patients were compliant.

In previous studies, satisfaction with care was found to be associated positively with older age among ambulatory care patients[18] and among patients treated for burns in emergency rooms.[19] In regard to compliance, no differences were found between the young and the elderly in the way they adhered to medical regimens.[20]

This study was an attempt to determine, first, if satisfaction with care differed in a group of young and elderly outpatients and, if so, which demographic and health variables significantly predicted levels of satisfaction for each age group. Finally, the relationship between satisfaction with care and compliance was determined for each age group. The hypotheses were that higher satisfaction with care would be found in older patients, that different predictors would be associated with satisfaction for the elderly as compared with the young, and that satisfaction in both age groups would be positively associated with compliance.

Method

A sample of 878 male patients attending ambulatory clinics in a Veterans Administration (VA) medical center were randomly selected during a three-year period. Patients were entered in the study if they had been known to the clinic for at least 12 months, and signed informed consent for interviews at baseline and two months later. Only 12 patients did not agree to participate. Information on age, race, social class (Hollingshead's two-factor index[21]), marital status, education and self-assessed health, measured on a 6-point scale ranging from very good to very poor, were obtained from the patients. Patients also completed the 45-item Satisfaction with Physicians and Primary Care Scale of Hulka et al.[22] Items are grouped into areas by content that measure satisfaction with the professional competence of physicians, personal qualities of physicians and cost/convenience of care. Although items can be grouped into these areas, most studies have reported the total score as a single measure of degree of satisfaction with care. Items are rated on 5-point scales, from strongly agree to strongly disagree, with some of the items phrased negatively and some positively. Higher scores represent a more negative attitude about care.

Twelve physicians provided care to the patients. They estimated severity of patients' conditions on a 6-point scale corresponding to that used by the patients for self-assessment. They also provided ratings on expectation in regard to patients' conditions two months later on a 3-point scale of better to worse.

Patients were followed up two months after they entered the study, and compliance with the medical regimen and other variables were assessed. In regard to measurement of compliance, all medications prescribed were documented from physician orders in the medical records and hospital pharmacy reports. Kind, amount and duration of prescriptions at the time of the index visit and for the subsequent two months were recorded. Nonmedication treatments, such as diets, special therapy or return visits to the clinic, were included for evaluation. At the home visit, the research nurse asked patients to recount what the physician had told them in regard to instructions for care. This was checked with data obtained from the records. Patients were asked if they adhered all, part or none of the time to each. They were then asked to show the nurse the bottles and containers of medications, and either a pill count or estimate of the amount of liquid in the bottle was made for comparison with the pharmacy record of what had been given to the patients. The information from the home visit, along with that from hospital records, was given to two physicians, who independently rated the degree of compliance on a 3-point scale, ranging from very compliant to very noncompliant. Physicians were blind to the age of the patients and other variables under study.

Also at the time of the home visit, patients were asked if they perceived their health to be better, the same or worse (3-point scale) than it had been two months previously. Patients also estimated their health status on a 6-point scale. Number of visits to the clinic or to non-VA physicians, or medications not prescribed by the VA were recorded for the two-month interval between baseline and follow-up visit.

Medical records of all patients were reviewed to determine the current number of diagnoses and medications, as well as number of clinic visits and hospitalization during the year preceding the date of intake into the study.

Data were analyzed by first comparing the level of satisfaction with care between patients aged 65 years and older (N = 267) and those under the age of 65 (N = 581) with a t-test between means. Satisfaction with care was then predicted separately for each age group by entering 12 variables collected at baseline into a stepwise multiple regression. Predictors were personal characteristics (race, education, social class and current marital status); distance from clinic; severity of illness (self-assessed health, number of diagnoses, number of medications, severity assessed by physicians and expectation of improvement as judged by the physicians); and utilization of care during the prior year (number of clinic visits and hospitalizations). The final step was to correlate level of satisfaction (baseline) with degree of compliance (two months later) for each age group separately.

Results

Satisfaction in the younger group was 119.6 (SD = 13.2), and that in the older group was 113.8 (SD = 12.4). Range for the scale was 45–225, with a lower score reflecting more satisfaction with care. Since these represent approximately midpoint values on a 5-point scale, it might be assumed that both groups were relatively satisfied with their overall care. In examining the particular areas of satisfaction, each group rated professional competence of physicians about the same (young = 37.20, elderly = 36.37), but the younger group was more negative than the older group on items related to personal qualities of physicians (37.42 v. 35.56) and cost/convenience of care (46.04 v. 41.82).

Table 1 shows the values for the items that were entered into the regression analyses predicting satisfaction with care. The average age of each group is also shown. In the younger group, the average age was 49 years, with a range of 23 to 65. The age of the older group averaged 69 years, with a range of 65 to 98. Approximately 13 per cent were black, which corresponds to the average in the state of

TABLE 1. Means and Percentages of the Variables Used to Predict Satisfaction With Care in Older and Younger Patients

Variables	Young (N = 581)		Older (N = 267)	
	Mean	Per Cent	Mean	Per Cent
Personal characteristics				
Age	49.1		69.3	
Black		11		05
Education (years)	12.1		10.8	
Social class	3.3		3.5	
Current marital status		80		75
Distance from hospital (miles)	21.4		19.4	
Severity of condition				
Self-assessed health	3.0		2.8	
Number of diagnoses	3.2		3.8	
Number of medications	3.6		4.0	
Severity (by MD)	3.7		3.8	
Expectation of improvement (by MD)	1.6		1.8	
Utilization				
Number of clinic visits prior year	4.8		5.4	
Number of hospitalizations	1.7		1.7	

Note: Self-assessed health rated 1–6, with higher worse; severity rated 1–6, with higher worse; and expectation of improvement rated as 1 = better to 3 = worse.

Florida. Most were middle class, and approximately three fourths were married. Patients lived about 20 miles from the clinic, perceived their health as fair to good, averaged approximately three diagnoses and three to four medications, and were estimated by the physicians to be more severely ill than they viewed themselves. On the whole, physicians expected these patients to be about the same or slightly worse in two months.

TABLE 2. Means of the Variables Collected at the Two-Month Follow-Up for Older and Younger Groups

Variables	Younger (N = 581)	Older (N = 267)
Number of additional visits	.02	.03
Number of hospital admissions	.00	.00
Number of medications	3.61	4.03
Patient evaluation of improvement	1.71	1.59
Self-assessed health	3.10	2.76
Compliance (by patient)	1.40	1.37
Compliance	1.52	1.48

Note: Patient evaluation of improvement was rated 1–3, from better to worse; self-assessed health was rated 1–6, with higher worse; and compliance was rated 1–3, with 1 being compliant all of the time, 2 part of the time, and 3 being mostly noncompliant.

All but 30 patients were followed up at two months. This represented a 3 per cent loss of follow-up data. Twenty-one patients (4 per cent) were in the younger, and nine (3 per cent) were in the older groups. The 30 patients lost to follow-up did not differ significantly on any of the demographic or satisfaction variables collected at baseline when they were compared statistically. Table 2 shows the follow-up data. It can be seen that a very small number of persons had been seen by the clinic during the two-month interval. Ambulatory care clinics generally schedule patients every three months for follow-up, unless otherwise indicated. This was one of the reasons a two-month follow-up time had been selected. No patients had been admitted to the hospital during the follow-up period. Medications were essentially the same as those that had been prescribed at baseline. No non-VA physicians had been consulted by the patients during the follow-up period. The patients' evaluation of their condition on a scale of 1–3, with 3 being the worst, was 1.7 for the young and 1.6 for the old. The correlation between the physician's expectation of patients' conditions two months later and the patients' ratings at that time was $r = 0.29$ for the young and $r = 0.08$ for the elderly. The correlation between patients' self-assessed health at baseline and two months later was relatively high ($r = 0.48$ for the young and $r = 0.59$ for the elderly). The correlation between self-assessed health and the physician's severity rating on the same scale at baseline was $r = 0.48$ for the young and $r = 0.26$ for the elderly. The mean score on compliance, using the patients' estimate, was 1.4 for the young and 1.4 for the elderly, with 1 being compliant all of the time and 3 none or almost none of the time. Average ratings of 1.52 and 1.48 were given by the physicians judging compliance for the two age groups. The agreement between the two physicians was $r = 0.72$. The difference between the elderly and the young was not statistically significant ($t = 0.34$).

Predicting Satisfaction with Care

Tables 3 and 4 show that three of the 12 variables significantly predicted satisfaction with care in each of the age groups.

As seen in Table 3, the best predictor of satisfaction in younger patients was severity of condition, as rated by the physician. Those who were more severely impaired were less satisfied with care. The next best predictor, in combination with severity, was miles from the hospital, which added to knowing severity at a 0.01 level of significance. The longer the distance from the clinic, the more was the dissatisfaction with care. The only other variable to add significantly to predicting satisfaction with care was number of clinic visits ($p < 0.05$), with more visits also associated with less satisfaction. These three variables accounted for 70 per cent of the predictable variance. All 12 predictors together accounted for 23 per cent of the total variance in satisfaction with care. Although several

TABLE 3. Predictors of Satisfaction With Care in 581 Patients Under the Age of 65

Predictor Items	r	Cumulative R^2	ΔR^2	Per Cent Predicted R^2
Severity rated by MD	0.34*	0.11	0.11	48
Distance from clinic	0.21*	0.14	0.03	61
Number of clinic visits	0.18*	0.16	0.02†	70
Current marital status	0.08†			
Education (years)	0.06			
Self-assessed health	0.13*			
Number of diagnoses	0.03			
Social class	0.14*			
Number of medications	0.04			
Black	0.09†			
Expectation of MD	0.04			
Number of hospitalizations	0.13*		0.07	
Using All Predictors		0.23		100

* $p < 0.01$.
† $p < 0.05$.
Note: Higher scores on satisfaction are worse, severity rated 1–6, with higher worse; self-assessed health 1–6, with higher worse; social class 1–5, higher = lower class; and expectation of physician 1–3, with higher worse.

of the other variables were significant at a zero-order level in their association with satisfaction, they did not add uniquely at a significant level to the cumulative variance.

Table 4 indicates that two of the same variables were also among the best predictors in the older group. Number of clinic visits was the best predictor and severity of illness, as judged by the physician, the third best predictor ($p< 0.05$). The second best predictor, however, was different from those identified in the young. The expectation of the physician as to whether the patient would improve or not in the next two months added significantly to number of clinic visits at a 0.01 level. Those patients whom the physician expected to be better were more satisfied with care. The interesting point is that even though there was probably overlap between physician expectation, number of clinic visits and severity, each measures something unique in itself in that each adds significantly to the other. Again, as with the younger group, there were other variables associated significantly with satisfaction at the zero-order level. They were, however, redundant variables and in combination with the three predictors did not add significantly to prediction. Among this older group, 77 per cent of the predictable variance was accounted for by three predictors. All 12 variables accounted for 22 per cent of the total variance in satisfaction with care among the elderly.

Correlation of Satisfaction with Compliance

The association between level of satisfaction at baseline and compliance two months later was $r= 0.41$ in the older group (significant at $p< 0.01$) and $r= 0.04$ in the younger group, which was not statistically significant. Table 5 shows how each of the areas of the scale related to compliance. None was associated with compliance in the young, and all were related to compliance in the elderly. In particular, the personal qualities of the physician and cost/convenience of care were more

TABLE 4. Predictors of Satisfaction With Care in 267 Patients Over the Age of 65

Predictor Items	r	Cumulative R^2	ΔR^2	Per Cent Predicted R^2
Number of clinic visits	0.32*	0.10	0.10*	45
Expectation of MD	0.28*	0.15	0.05*	68
Severity by MD	0.16*	0.17	0.02†	77
Distance from hospital	0.11†			
Number of medications	0.13†			
Current marital status	−0.13†			
Number of diagnoses	0.18*			
Black	−0.08			
Social class	0.05			
Education	0.03			
Self-assessed health	0.07			
Number of hospitalizations	0.12†		0.05	
Using All Predictors		0.22		100

* $p < 0.01$.
† $p < 0.05$.
Note: Higher scores on satisfaction are worse, severity rated 1–6, with higher worse; self-assessed health 1–6, with higher worse; social class 1–5, higher = lower class; and expectation of physician 1–3, with higher worse.

closely related to compliance in the older group than was the professional competence of the physician.

Discussion

The hypothesis that the older patients would be more satisfied with care was supported. The finding that the younger patients were more dissatisfied in regard to items that measured the convenience of care is somewhat understandable. Many in this group were employed and may have had long waits when they came for appointments. They also lived an average of a few miles farther away from the clinic than did the older group. Elderly patients are perhaps less likely to complain about long waits and sometimes even see the clinic visits as a social event. Furthermore, the elderly as a group may be less likely to report problems. For example, they are known to overestimate favorable health in comparison with objective health measures,[23] and the same could be true for their satisfaction-with-care ratings.

The hypothesis that there would be different predictors for satisfaction in the two age groups was only partially supported. Two variables, clinic visits and severity, as judged by the physician, were predictors for both age groups (although differing in order of importance). The greater the severity estimates by the physician, the more dissatisfaction was found in each age group. However, severity accounted for most of the variance in satisfaction in the young but was the third best predictor in the old. Severity of condition has been found to be associated with satisfaction in other studies,[6,8] in a similar way. The fact that it was more associated with satisfaction in the young than in the elderly in this study may indicate that when age is taken into account, older patients are more likely to accept some chronic conditions as a natural part of aging. Number of clinic visits was the best predictor of satisfaction in the elderly, with more visits associated with less satisfaction. This same variable describing clinic visits was the third best predictor in the young. While it can be concluded that clinic visits and severity add to each other in prediction, indicating that each measures something unique, it is also likely that there is considerable overlap between the two in regard to both being measures of severity as viewed by the physician who schedules appointments.

TABLE 5. Correlations Between Satisfaction With Care and Later Compliance for Two Age Groups

Satisfaction Variables	Correlations With Compliance	
	Young	Elderly
Professional competence of MDs	0.02	0.07*
Personal qualities of MDs	0.01	0.22†
Cost/convenience of care	0.05	0.44†
Overall satisfaction	0.04	0.41†

* $p < 0.05$.
† $p < 0.01$.
Note: Higher scores on satisfaction are less satisfied, and higher scores on compliance are less compliant.

The predictors that differed between groups were distance from the clinic in the young and physician expectations of improvement in the old. As mentioned, distance may be particularly important to the younger patients if they must take time off from work, and it is likely to influence their ratings of satisfaction when convenience with care is evaluated. The expectation of the physician regarding prognosis helped to explain satisfaction with care in the elderly. This can perhaps be explained by looking at the combination of predictors in each group. In the young, it was more severity, longer distance from the hospital and more clinic visits that were associated with dissatisfaction. In the elderly, it was more frequent visits, less expectation for improvement and more severity that predicted dissatisfaction. One could argue that this combination in the elderly, all influenced by the physician, represented "more severity" as viewed by the physician. At the same time, however, the correlation between how the physicians viewed the patients in regard to severity and how the patients viewed themselves was not as highly correlated in the elderly as in the young. Thus, the older patients rated their health as better than did the physicians. One might ask which was the more valid assessment. Generally, it is assumed that the physician's assessment is the objective measure. Therefore, in most studies[18,23] the elderly have been found to overestimate their health as favorable in comparison with physician estimates. Conversely, how much the physician's estimate is influenced by his/her views on aging, particularly in regard to some stereotype of the elderly as frail and weak, is not known. The issue here is not so much which view was right, but rather that patient and physician views were incongruent for some of the elderly.

It seems obvious that the patient and physician need to share some common understanding about the person's condition and treatment. With the elderly, however, this may represent a dilemma. It has been shown in other studies[23,24] that older peoples' assessment of their health was highly related to their functional status. Some persons with poor objective health ratings perceived their health as good and continued to function adequately in everyday activities of living, while others who saw their health as poor, without any objective evidence of this severity, did poorly in physical functioning. The relevant question would appear to be that of how physicians and patients can get together on their views, so that treatment starts from a common base. This takes on added significance in light of other studies, in which lack of congruence between physician and patient, as well as severity of the patient's condition, were found to be associated with compliance with the medical regimen. The health belief model described by Rosenstock[25] and elaborated upon by Becker[12] suggests that patients must perceive their condition as serious and treatment to be of potential help in order for compliance to be assured. However, the health belief model deals more with preventive health care than with chronic, ongoing health problems, although some would suggest that the same model applies for both kinds of care.

If the patient is made to see his/her health as poor, whether this information is transmitted directly or indirectly by the physician, compliance might be increased, but possibly at the expense of a self-fulfilling prophecy. The older person would have to change his/her self-image from that of someone in good to someone in poor health, and as a result might function less well (unless the treatments prescribed were indeed a powerful remedy).

On the other hand, physicians cannot be expected to change their views about the severity of the condition, if they have made their prognosis on the basis of individualizing the patient and not on a stereotype of the elderly as a group. Most of the elderly under care have multiple chronic conditions. Furthermore, they live with these conditions for a considerable time. Cure is not usually the objective for these patients, as compared with younger patients with more acute illnesses. It is expected that physicians would become frustrated that treatments are not always available to improve the patients' conditions significantly. With the elderly, having someone who understands their condition, someone on whom they can rely, and someone to whom they can turn when they perceive their health worsening, is very important. This requires frequent visits of short duration, during which the "caring" role of the physician is uppermost. Medications are sometimes then adjuncts to maintaining functional status.

Finally, the differential effect of satisfaction on compliance by age was somewhat unanticipated. It had been hypothesized that satisfaction would be related to compliance significantly in both age groups, but only in the older group was a significant association between the two found. Age groups did not differ in regard to level of compliance, with noncompliance being present in each age group. It should be recognized that the level of association between satisfaction and compliance that was found in the elderly could only be considered moderate. In other words, other factors, not measured by satisfaction with care, still predict a large proportion of the variance in compliance. However, in the younger group no relationship between the two existed, at least in terms of the variables used in the study to measure satisfaction. It is tempting to speculate that since the physician's expectation influenced satisfaction in the elderly, when the physician did not believe that his/her contact with the patient was helpful, somehow the patient sensed this in the physician's attitude and thus put little faith in the medical regimen as well.

In summary, this study highlights some of the variables that impact on satisfaction with medical care in young and elderly, and suggests that in the elderly, satisfaction influenced compliance with the medical regimen. Findings also suggest that for older patients there may be value in the physician's assuming more of a "caring" then "curing" role in expectation and prognosis. This, in turn, might lead to a reduction in rates of noncompliance for elderly chronically ill patients.

References

1. Roghmann KJ, Hengst A, Zastowny TR. Satisfaction with medical care: its measurement and relation to utilization. Med Care 1979;17:461.

2. Cafferata GL, Roghmann KJ. Patient satisfaction and the use of health services: explorations in causality. Paper read at the Eastern Sociological Meetings, 1978.

3. Starfield B. Health services research: a working model. N Engl J Med 1973;289:132.

4. Koos EL. The Health of Regionville. New York: Hafner, 1954.

5. Linn MW, Linn BS. Qualities of institutional care that affect outcome. Aged Care and Services Review 1980;2:1.

6. Gerst A, Rogson L, Hetherington R. Patterns of satisfaction with health plan coverage. Inquiry 1969;6:37.

7. Hulka BS, Kupper LL, Daly MB, et al. Correlates of satisfaction and dissatisfaction with medical care: a community perspective. Med Care 1975;13:648.

8. Francis V, Korsch BM, Morris MJ. Gaps in doctor–patient communication: patients' response to medical advice. N Engl J Med 1969;280:535.

9. Fisher AW. Patients' evaluation of outpatient medical care. J Med Educ 1971;46:238.

10. Davis M. Physiologic, psychological and demographic factors in patient compliance with doctors' orders. Med Care 1968;6:115.

11. Blum RH. The Management of the Doctor—Patient Relationship. New York: McGraw Hill, 1960.

12. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care 1975;13:10.

13. Stimson GV. Obeying doctor's orders: a view from the other side. Soc Sci Med 1974;8:97.

14. Davis M. Variations in patients' compliance with doctors' advice: an empirical analysis of patterns of communication. Am J Public Health 1968;58:274.

15. Gordis L, Markowitz M, Lilienfeld AM. The inaccuracy in using interviews to estimate patient reliability in taking medications at home. Med Care 1969;7:49.

16. Blackwell B. Drug therapy: patient compliance. N Engl J Med 1973;289:249.

17. Kasl SA, Cobb S. Health behavior, illness behavior, and sick role behavior. I. Health and illness behavior. Arch Environ Health 1966;12:246.

18. Linn BS, Linn MW, Knopka F. The very old patient in ambulatory care. Med Care 1978;16:604.

19. Linn BS. Continuing medical education: impact on emergency room burn care. JAMA 1980;55:512.

20. Linn BS, Linn MW. Ambulatory care and the elderly. In: Pascarelli EF, ed. Principles of Hospital Based Ambulatory Care. New York: Appleton-Century-Croft (in press).

21. Hollingshead AB. Two Factor Index of Social Position. New Haven, 1957. (Mimeograph.)

22. Hulka BS, Zyzanski SJ, Cassel JC, Thomson SJ. Scale for the measurement of attitudes towards physicians and primary medical care. Med Care 1970;8:429.

23. Linn BS, Linn MW. Objective and self-assessed health in the old and very old. Soc Sci Med 1980;14:311.

24. Linn MW, Hunter KI, Linn BS. Self-assessed health, impairment, and disability in Anglo, black, and Cuban elderly. Med Care 1980;18:282.

25. Rosenstock IM. What research in motivation suggests for public health. Am J Public Health 1960;50:295.

MEDICAL CARE
June 1982, Vol. XX, No. 6

Participants in Noncompliance Research: Compliant or Noncompliant?

JANICE A. RAMSAY, PH.D.

Patients have the option to participate in many investigations of noncompliance for medical recommendations. The question can therefore be raised about whether a tendency to volunteer for medical research is consistent with a tendency to comply with other medical treatment requests. In short, are volunteers or participants generally compliant relative to nonparticipants? This study examined past compliance for recommended follow-up appointments in patients who were requested to participate in a blood pressure assessment program and in a subsequent study on compliance for hypertensive regimens. Hospital records of the patients were reviewed for the two-year period that preceded the request to participate. Analysis of variance demonstrated significant differences in past compliance for participants and nonparticipants. In general, participants were compliant compared with nonparticipants. The results of this study suggest that investigations of noncompliance have used samples that are select for compliant behavior. The implications of this study on compliance research are discussed.

SEVERAL REVIEWS of the literature have criticized sample selection methods in research on compliance with medical regimens.[1,2] However, investigations to support or refute those criticisms have not been conducted. This article first defines one area of sampling bias in compliance research and then reports an empirical investigation of the problem.

The general concern over sampling methods has been the degree to which the sample is representative of the population. Researchers usually obtain samples from clinical settings that are easily accessible, such as hospital outpatient departments or private medical practices. Patients in these settings may have characteristics that distinguish them from patients who are in other clinical settings and from patients who, although symptomatic, have not sought health care at all. Furthermore, patients who drop out of treatment before the sample is selected are likely to have characteristics that distinguish them from other patients.

In addition, since volunteers and nonvolunteers are known to have different characteristics,[3] the extent to which the sample represents the population is further compromised when patients have the option to participate in studies on compliance. One can only guess at the effects on measures of compliance when volunteer samples are used. One possibility is that patients who volunteer may do so because they are committed to the treatment under investigation and may be more likely to comply with the treatment recommendations. A second possiblity is that patients who agree to participate in a study do so because of a tendency to be com-

---

From the Department of Social and Preventive Medicine, University of Manitoba, Winnipeg, Manitoba.

Research for this study was supported by a grant from the National Health Research and Development Programs, Health and Welfare Canada (Grant No. 6607-1103-43).

0025-7079/82/0505/0615/$00.90 © J. B. Lippincott Co.

pliant, a tendency that may carry over to requests related to the treatment. If the sample consists of compliant individuals, investigation of noncompliance may be problematic.

Few studies on noncompliance for medical recommendations have directly addressed questions related to sample selection biases. Typically, comparisons are made among participants and nonparticipants on sociodemographic or psychological factors rather than on adherence to the regimen.[4–6] However, one study has investigated differential compliance rates for participants and nonparticipants. In that investigation, psychiatric outpatients were contacted and asked to participate in a study done at a regularly scheduled appointment and taking several hours.[7] The outcome measure of volunteer status was compared with psychological and sociodemographic data obtained at earlier appointments. Patient hospital records, obtained some time after the study, were also examined. It was found that 57 per cent of those who had volunteered had continued therapy or been referred elsewhere. On the other hand, 69 per cent of nonvolunteers had terminated early from the program.

The study is important since it suggests that nonparticipants in a study within a psychiatric setting also tend to be noncompliant. However, another interpretation of the data is possible. Negative experiences in the psychiatric clinic prior to the request may have precipitated both the refusal to participate in the psychotherapy study and the decision to drop out of therapy. Therefore, the compliance of the patient may not have been a factor in early withdrawal.

The literature, then, does not contain evidence to support or refute a view that participants in medical investigations are compliant. The purpose of the present study is to fill that void. It was hypothesized that compliance with medical recommendations made prior to a request to participate in a clinical study would be lower for nonparticipants than for participants.

Method

Sample

Patients investigated in the present study had attended the Health Sciences Centre, Winnipeg, Canada, a tertiary care hospital located in the inner city. The hospital serves the entire province, but the ambulatory care departments (Primary Health Care, Outpatient Department and the Emergency Department) are used primarily by the population from nearby areas. Patients attending the ambulatory care departments are often from a lower socioeconomic level. In addition, the ambulatory departments are frequently utilized as a source of medical care by North American Indians (natives), who constitute a major ethnic group in the area.

A blood pressure screening program was conducted in the ambulatory care departments during the period of March 1978 to April 1979. Patients with recorded diastolic blood pressures of 100 mmHg or greater were invited to have further blood pressure assessment and were informed of the research nature of the program. After assessment, if the patients were found to require treatment, they were invited to receive care in a hypertension clinic in which compliance was to be investigated.

Five groups of participants and nonparticipants were identified on the basis of three criteria: responses to two requests to participate and hypertension status as determined by resting blood pressures taken during assessment (Table 1). The five groups were classified as follows.

1. Unknown-nonparticipant: patients who would not participate in blood pressure assessments and whose health status and need for treatment were therefore unknown.
2. Hypertensive–nonparticipant: patients who participated in blood pressure assessment and were found to require

TABLE 1. Criteria for Participant and Nonparticipant Groups

Group	Elevated Screening Blood Pressure	Response to Request to Be Assessed	Resting Blood Pressure Elevated	Response to Request to Participate in Compliance Study	Attended Compliance Study	N
Unknown–nonparticipant	yes	no	?	—	—	40
Normotensive–participant	yes	yes	no	—	—	40
Hypertensive–participant	yes	yes	yes	yes	yes	40
Hypertensive–nonattender	yes	yes	yes	yes	no	13
Hypertensive–nonparticipant						
Completed screening	yes	yes	yes	no	—	9
Did not complete screening	yes	yes	yes	no	—	25

treatment for hypertension, but refused follow-up.

3. Normotensive–participant: patients who agreed to participate in blood pressure assessment, but were found to be normotensive and therefore did not require follow-up.

4. Hypertensive–participant: patients who participated in blood pressure assessment, were found to require treatment for hypertension, agreed to participate in follow-up and did attend.

5. Hypertensive–nonattender: patients who participated in blood pressure assessment, were found to require treatment and agreed to participate in a follow-up progrm, but then did not attend.

There were 741 suitable patients identified through the screening program and distributed over the five groups as follows: unknown–nonparticipant (N = 259), hypertensive–nonparticipant (N = 80), normotensive–participant (N = 268), hypertensive–participant (N = 115) and hypertensive–nonattender (N = 19). The sample, in this study, was randomly drawn from these 741 patients.

Procedure

Lack of consistency in compliant behavior across treatment recommendations[8,9] determines that comparisons among patients should be made on the same compliance measure. The measure likely to be available for the largest number of patients was attendance, and it, therefore, was chosen for this study.

In this study, attendance was investigated for the time *prior* to the assessment request in order to control for the possibility that the request influenced later attendance for recommended appointments. The assessment program was separate from the ambulatory care departments, thus reducing the chance that prior experiences in the departments would influence decision to participate in assessment.

Two years of visits to the ambulatory care departments were reviewed for each patient. Since not all patients in the five groups had previously been requested to make return hospital appointments, information on the critical measure was missing in some cases. A minimum of three instances of physician-recommended return visits were necessary to admit a patient to the study. Thus, an ineligible patient would be one having fewer than three physician recommendations for follow-up in the two years prior to blood pressure assessment. The charts were randomly ordered and sequentially reviewed by groups until 40 cases per group were found. The exception was the hypertensive–nonattender group, with only 19 members, all of whom were included.

Each medical visit noted in the chart in the two-year period was reviewed for: compliance with each follow-up recommendation; address and employment status (recorded at the visit that precipitated referral to the assessment program); date of birth; gender; and ethnic group (native or non-native). This information was obtained for all patients, whether or not they met the criterion of three follow-up visits. Data on patients not meeting the attendance criterion (ineligible patients) were later analyzed to determine selection factors within this study.

For the purpose of this study, recommendations for follow-up visits consisted only of recommendations for visits by and with medical practitioners, since a restriction of this type ensured uniformity in the study. Therefore, visits for laboratory or clinical investigations and visits with nurses, dieticians, dentists and physiotherapists were excluded.

Some of the data that were collected require explanation. Attendance compliance was expressed as a ratio: total appointments kept to total appointments recommend over the two-year review period. Address provided the only consistent indication of socioeconomic status found in the chart, and this was coded using specific geographic borders defining: 1) the core area of the city; 2) regions adjacent to the core area; and 3) all other regions.

Employment status, routinely noted on admission records of nonscheduled ambulatory care visits, was obtained from the record for the date of the assessment contact. Ethnic status was not always noted by those making entries in health records, but patients who were North American Indian were often labelled. Therefore, the records of the patients were searched for any reference to ethnic status. If none was found, the patient was categorized as non-native.

Results

Cases Reviewed

In total, 393 patient charts were reviewed. The required 40 patients per group who met the three-visit criterion were selected in the following groups: unknown–nonparticipant, normotensive–participant and hypertensive–participant. All patient charts for hypertensive–nonparticipants and hypertensive–nonattenders were reviewed, but only 34 and 13, respectively, met the selection criteria and were included in the study.

In the course of the review, it was noted that the hypertensive–nonparticipants could be subdivided into two groups. One group of patients preferred to return to private physicians for treatment of their hypertension (completers, N = 9). The other group consisted of patients who had failed to complete assessment or had declined treatment of hypertension without mentioning that they would consult a private physician (noncompleters, N = 25).

Some descriptive statistics suggested that there might be substantial differences between the two subgroups of the hypertensive–nonparticipants. Rates of alcoholism, blood pressure and age were not alike (Table 2). Therefore, it was decided to subdivide the group.

Since early analysis had given evidence of noncomparability of the two subgroups of hypertensive–nonparticipants, it became apparent that all data analysis for hypothesis testing would have to be accomplished with six groups of markedly unequal sizes. Both parametric and nonparametric analyses were employed (Analysis of Variance, Chi-Square and Irwin-Fisher test).

Ineligible Patients

Among the 393 charts reviewed, 226 did not meet the criterion of three recommended follow-up visits and were ineligible for the study. However, demographic and health-related data were obtained on the ineligibles to determine the nature of selection bias in this investigation.

Since a criterion of three recommended follow-up visits was employed, the ineligible patients differed from those investi-

TABLE 2. Characteristics of Hypertensive–Nonparticipant Completers and Noncompleters

Characteristic	Hypertensive–Nonparticipants: Completers	Hypertensive–Nonparticipants: Noncompleters
Rate of alcoholism	11%	32%
Age		
Mean	52.7 years	41.8 years
Standard deviation	8.9	14.2
Screening diastolic blood pressure		
Mean	104.4 mmHg	107.6 mmHg
Standard deviation	5.1	6.9

gated in terms of total ambulatory visits and total recommended follow-up visits. Ninety-six per cent of ineligible patients had six or fewer ambulatory visits (28 per cent of eligible patients had as few visits). Of the 226 who were ineligible, 138 had no recommended follow-up visits, 67 had one recommended visit, and 21 had two recommended visits. The charts of these patients contained inadequate information on attendance compliance.

Besides medical visits, there were many other ways in which the eligible and ineligible patients differed. The patients from the two eligibility classes were compared on blood pressure, past history of hypertension, social problems, employment, ethnic group, gender and place of residence (Table 3).

Place of residence was significantly different for the two classes. Eligible patients lived in the core area more frequently than expected, whereas ineligible patients more frequently resided in areas beyond the core area. Patients admitted to the study probably were of a lower socioeconomic status than patients not admitted.

During the chart review, various psychosocial and environmental problems were noted in physicians' records. It was acknowledged that there are deficiencies in using recorded social problems because not all problems are noted by physicians, and those that are recorded may represent errors in documentation or biases of the physician. However, the decision was made to use the information that was recorded while keeping in mind the potential bias when interpreting the data. Patients in the two eligibility classes were compared for the following categories of social problems: no problems noted, alcoholism and other problems. The category "other problems" was a combination of a number of low-frequency categories (involvement with the law, family and financial problems). Ineligible patients were less likely to have social problems noted in their charts than had been expected. The opposite was true for eligible patients.

The Irwin-Fisher test was used to compare the two eligibility classes on employment status. Ineligible patients were more likely to be employed. However, patients who were admitted to the study demonstrated no difference between employment and unemployment rates. As for ethnic group, ineligible patients were more likely to be non-native. Gender also was significantly different, with more males among ineligible patients.

No difference was found between eligible and ineligible patients for probability of blood pressure falling above or below the grand median of 102 mmHg. However, there was a significant difference in rates of documentation of past history of hypertension. Ineligible patients were less likely to have a past history of hypertension noted in their hospital records.

TABLE 3. Characteristics of Eligible and Ineligible Patients in the Study

Characteristic	Ineligible (N = 226)	Eligible (N = 167)	Statistic
Gender			
Male	144 (64%)	89 (53%)	Z = 6.58
Female	82 (36%)	78 (47%)	$p < 0.05$
Residence			
Core	70 (31%)	83 (50%)	$\chi^2 = 27.08$
Fringe	63 (28%)	55 (33%)	df = 2
Other	93 (41%)	29 (17%)	$p < 0.005$
Social problems			
None noted	153 (68%)	66 (40%)	$\chi^2 = 31.40$
Alcoholism	50 (22%)	64 (38%)	df = 2
Other	23 (10%)	37 (22%)	$p < 0.005$
Ethnic group			
Native	23 (10%)	70 (42%)	Z = 7.27
Non-native	203 (90%)	97 (58%)	$p < 0.05$
Employment rate	143 (63%)	72 (43%)	Z = 3.96 $p < 0.05$
Past history of hypertension	32 (14%)	55 (33%)	Z = 4.41 $p < 0.05$

Chi-square analyses done separately on eligible and ineligible patients, for participant and nonparticipant groups by ethnic status, place of residence, employment status and social problems, demonstrated no significant participant group differences. In addition, there were no participant–nonparticipant group differences for ineligible patients on documented past history of hypertension. This was not so, however, for some groups of patients admitted to the study: normotensive–participants had lower frequencies of previously documented hypertension, and hypertensive–participants had higher frequencies than expected.

Attendance

The attendance compliance ratio was determined by dividing the total attended recommended follow-up visits over the two-year review period by total recommendations for follow-up. Total recommended visits ranged from three to 69 (mean 7.7 visits, standard deviation 7.5). There were no significant group differences in frequency of medically recommended visits or total ambulatory visits (recommended visits and visits initiated by the patient). Further, the simple correlations of the attendance compliance ratio with either total ambulatory visits or total recommended visits were near zero (total visits, $r = 0.06$; total recommended visits, $r = 0.09$).

Since the Cochran and Bartlett-Box tests demonstrated homogeneity of variance for the attendance compliance ratio with groups of unequal sizes, analysis of variance was used to determine that there were significant group differences. Table 4 shows the summary table for the analysis of variance with group means and standard deviations for attendance compliance ratios.

Discussion

Sample Selection Bias

In this study it was proposed that prior compliance with recommendations about follow-up appointments would be greater among participants in a study on noncompliance than among nonparticipants.

TABLE 4. Analysis of Variance Summary Table
Attendance Compliance by Group

Source	DF	Sum of Squares	Mean Squares	F	Probability
Between groups	5	0.8163	0.1633	2.33	<0.05
Within groups	161	11.2988	0.0702		
Total	166	12.1151			

Group	N	Mean Attendance Compliance	Standard Deviation
Hypertensive–nonparticipant noncompleters	25	0.5578	0.3019
Unknown–nonparticipant	40	0.6116	0.2523
Hypertensive–nonattender	13	0.6410	0.2819
Hypertensive–participant	40	0.6750	0.2805
Normotensive–participant	40	0.7546	0.2253
Hypertensive–nonparticipant completers	9	0.7611	0.2804

The data indicate that different rates of attendance compliance exist for participants and nonparticipants. Participants fell in the upper range of attendance compliance, and, in general, nonparticipants had past attendance compliance rates in the lower range.

The interpretation of the finding in this study of a biased sample depends in large part upon the extent to which one is willing to assume that attendance compliance predicts compliance for other medical recommendations. Based on the literature, compliance for one recommendation does not predict behavior for other recommendations. Two points argue against the literature. One is that the patient who does not attend will not have the opportunity to receive the medical recommendations associated with the study: nonattendance is therefore maximum noncompliance in this situation. The second is that there is reason to question earlier findings of low correlations among different regimens on the basis of demonstrated biased samples found in this study. Reported weak correlation coefficients between any two regimens may be a statistical consequence of using a restricted range of compliant behavior (as found among participants). The relationship may be stronger when data from the nonparticipants are included.

Therefore, the present study interprets the results to indicate that participants in an investigation of medical compliance are generally more compliant than those who choose not to participate. As a consequence, participants may have a tendency to comply with other medical recommendations, a tendency that may not be evident among nonparticipants.

Research outcomes of past studies may not apply to nonparticipants. Rates of noncompliance or factors associated with noncompliance may be underestimated. Various strategies or techniques that have been investigated to determine the effect on compliance may have a different effect on those not included in the study.

Several points should be made about this study. First, attendance compliance of the hypertensive–nonparticipant completers was in the upper range and, therefore, appears to deviate from the overall pattern. Patients in this group declined treatment in the Hypertension Clinic and indicated that they would seek treatment from another source. Their nonacceptance of treatment in the Hypertension Clinic may reflect a preference for an alternative treatment facility rather than noncompliant behavior per se, and further investigation is warranted. However, information was not available about whether treatment was actually sought.

It should be recognized that the present study has its own problems with sample selection bias, problems that tend to be common in other investigations. First, the population chosen for the study on compliance was narrow; it included individuals from one region, who were using hospital ambulatory facilities. Second, the study was select for disease; only hypertension was investigated. Finally, in order to find sufficient data on past attendance compliance in the present study, some patient records were excluded. From the comparison of eligible and ineligible patients, those included in this study came from a lower socioeconomic group and made greater use of health services.

However, differences in use of health services may have resulted from socioeconomic differences. The ineligible patients with their higher socioeconomic status may have sought most health care at suburban hospitals or in private physicians' offices. The hospital records would then contain less information about past medical history and social history, suggesting fewer health and social problems than may exist.

Regardless of the existence of differences in use of health services, the patients admitted to the study and the ineligible patients seem to differ on socioeconomic status, which may or may not be a factor in compliant behavior. Thus, any interpretation of findings in this study related to broad indicators of compliance should consider that our sample is socioeconomically select. However, as other compliance investigations have been conducted with similar samples, the results on sample bias can be generalized to other studies.

Conclusions

The purpose of this investigation was to determine if samples chosen for research on compliance consist of those who have a proclivity to comply. It was found that past compliance for recommended medical visits did vary across groups of participants and nonparticipants. Those who ultimately participated in a compliance study had past attendance compliance rates falling in the upper range. The results of investigations using similar volunteer samples should be generalized only to those who would participate in a compliance study.

Acknowledgments

The author wishes to thank David G. Fish, John S. McIntyre, Neena Chappell, Linda Henderson and Audrey Blandford for their contributions.

References

1. Sackett DL. The magnitude of compliance and noncompliance. In: Sackett DL, Haynes BR, eds. Compliance with therapeutic regimens. Baltimore: John Hopkins University Press, 1976:9.
2. Ramsay JA. Sample selection bias in investigations of medical treatment noncompliance. Doctoral dissertation. Winnipeg: University of Manitoba, 1981.
3. Rosnow RL, Rosenthal R. The volunteer subject revisited. Australian Journal of Psychology 1976;28:97.
4. Gould RL, Paulson I, Daniels-Epps L. Patients who flirt with treatment: the silent population. Am J Psychiatry 1970;127:524.
5. Slaikeu KA, Tulkin SR, Speer D. Process and outcome in the evaluation of telephone counseling referrals. J Consult Clin Psychol 1975;43:700.
6. Tilson E. Organization for long term management of hypertension: the trade union as a compliance mechanism in the treatment of hypertension. Bull N Y Acad Med 1976;52:714.
7. Carr JE, Whittenbaugh JA. Volunteer and nonvolunteer characteristics in an outpatient population. J Abnor Psychol 1968;73:16.
8. Donabedian A, Rosenfeld LS. Follow-up study of chronically ill patients discharged from hospital. J Chronic Dis 1964;17:847.
9. Nelson EC, Stason WB, Neutra RR, et al. Impact of patient perceptions on compliance with treatment for hypertension. Med Care 1978;16:893.

MEDICAL CARE
June 1982, Vol. XX No. 6

Health Diaries: Strategies for Compliance and Relation to Other Measures

GEOFFREY R. NORMAN, PH.D., ALLAN H. MCFARLANE, M.D.,
DAVID L. STREINER, PH.D., AND K. NEALE

This article describes experience with the use of health diaries in a two-year prospective study of the influence of the psychosocial environment on the health status of 500 subjects in the Hamilton area. Three strategies were used to maintain compliance: 1) random sampling of three days within each two-week interval; 2) a lottery ticket incentive; and 3) telephone follow-up. The results of an initial telephone and lottery ticket study demonstrated a significant effect on compliance. The data from the two-year study were analyzed to characterize health change based on diary reports over successive six-month intervals and to determine the relationship of reported health to medical utilization data. The results indicated that subjects reported symptoms on about one third of the days surveyed but took no action on about one third of these symptom days. Only one quarter of the symptom days resulted in a change in usual activity, and fewer than one tenth of the symptom days resulted in time off from work or physician visits. The diary variables showed a low, positive correlation with health utilization.

ANY ATTEMPT TO STUDY the factors that influence health and disease in a general population must deal with the issue of measuring episodes of illness. The most readily accessible measures of illness—hospital or ambulatory records—represent only a small fraction of the illness actually experienced by individuals, as pointed out by White[1] in his classic paper. Utilization data can be supplemented by personal interviews, but the cost of such an approach precludes its use on a regular basis, as does the finding that subjects have limited recall for health problems over a time interval of weeks or months.[2]

One alternative that has been used to a limited degree for several decades is the health diary—a personal, day-by-day record of health-related events completed by patients over a period of weeks or months. Verbrugge[3] has recently reviewed studies using health diaries reported in the literature from 1940 to the present. Of the 19 studies cited, only seven used the diary as the primary source of health data. The remainder used the diary either as a methodological check on other measures of health or as a reminder during subsequent interviews. Two of the seven studies, both conducted in the early 1950s, used diaries over an extended time—from one to ten years. None of the recent studies has attempted to extend the diary period beyond four to six weeks.

There appears to be good reason to limit the total duration of diary usage. Although initial compliance appears to be relatively

---

From the Department of Psychiatry, McMaster University, Hamilton, Ontario.

This research was supported by Health and Welfare Canada under grant NH-606-1420-42.

Reprint requests: Dr. G. R. Norman, Room 2V10, McMaster University Medical Centre, Hamilton, Ontario, Canada.

0025-7079/82/0505/0623/$00.85 © J. B. Lippincott Co.

high among those who consent to participate, there is a consistent drop in the reporting of illness with the passage of time, amounting from 5–25 per cent over a three-month period. Mooney[4] has attributed this to "fatigue," a lack of willingness to complete the diaries in the same detail as time passes.

Thus, although the diary has some very attractive features, principally its ability to tap into mild or acute episodes of illness that would not be reported to a physician, it has been used infrequently in longitudinal studies, apparently because of a decline in both participation and completeness over time.

Methods and Materials

The present article reports on the use of a diary as one source of health data from a large sample of patients in a longitudinal study to assess the influence of the psychosocial environment on health status.

This prospective study involved a cohort of 518 adult patients, randomly sampled from ten family practices in the Hamilton region. Subjects were followed over a two-year period, using a variety of data sources including a home interview completed upon entry into the study and at six-month intervals, documentation of physician contacts using an encounter form and the health diary, completed for three days every two weeks.

The study had two hypotheses:

1. The positive relationship between recent stressful events and illness will be modified by personal coping resources. Effective coping resources will weaken the relationship, whereas ineffective personal coping will strengthen it.
2. The positive relationship between recent stressful events and illness will be modified by support from the social network. The stronger such support, the weaker will be the relationship.

To test the hypotheses a more sensitive measure of change in health than physician visits was required, in view of the fact that this population had approximately one visit per subject in a six-month interval. For this reason, a health diary, allowing sampling of health status at two-week intervals, was included in the study.

The personal interview contained a number of standard scales: the Recent Life Change Questionnaire of Holmes and Rahe,[5] Rotter's Locus of Control Scale,[6] the Multidimensional Health Locus of Control,[7] a scale of health values,[8] a scale developed by the investigators to assess support available from social relationships,[9] and Langner's scale of physiological and psychological symptoms.[10]

The design and methods of the study are described in more detail elsewhere.[11] The purpose of the present article is to examine the use of the health diary: methods of administration, data on compliance and reported symptoms, and relationships to other measures of illness.

Methods of Administration

At the time of the initial home interview, subjects were informed about the diary by the interviewer. The first set of diaries was mailed out within two weeks, and thereafter a set of diaries was mailed out every two weeks to be completed for three consecutive days, selected at random by the research team. The choice of a random three-day period was made for a number of reasons:

1. By confining the diary period to a three-day sample every two weeks, we felt that we could obtain adequate documentation of the change in the subject's health over the duration of the study without major inconvenience to the subjects.
2. The random selection of the diary days insured that the sampled days would be a random sample of individual health and would also provide a small degree of novelty. The sampling of days was conducted so that for each individual, and for the sample, each day of the week was sampled uniformly.
3. Subjects were asked to complete the diary for the three days at one sitting, on the last day of the diary period. We felt this would reduce the work load of subjects, but

would not seriously affect their ability to recall information.

Within 48 hours of the scheduled day of completion of the diary, subjects were contacted by phone and asked if they had completed and returned the diaries. The telephone follow-up was conducted for all subjects.

Finally, as an incentive, a provincial lottery ticket, costing one dollar, was mailed to subjects after every three diary periods (six weeks), regardless of whether diaries had been completed or not. Ideally, we would have preferred to have telephone follow-up only when diaries were not returned, as the telephone calls became a minor irritation to some subjects, but this was not possible because of logistics. Similarly, it would have been preferable to send a lottery ticket only for completed diaries, but this would have added considerably to the logistical problems of the study.

Information about each day in the diary period was documented on a separate page. The variables assessed included reported symptoms and their subjective severity rating, actions taken (time off from work, health professional visit, prescription or over-the-counter medication, etc.), and an overall rating of health on a five-point scale. The diaries were sent with a stamped, self-addressed envelope for return mailing.

The data were keypunched and entered into a summary patient file through an input program that developed a number of summary variables over each six-month period between interviews.

Diary Compliance—Pilot Study Results

As already described, two incentives were incorporated into the diary administration in this study: telephone encouragement at the end of each diary period and a lottery ticket sent to each individual at the end of every three diary periods. Because these techniques add to the cost of acquiring the diary data, they were initially tested on a pilot sample of subjects (n = 117) drawn six months prior to the commencement of the main study. The 117 patients were randomly assigned to six groups of approximately equal size. Three groups received a lottery ticket every two months, and three did not. Crossed with this factor, two groups received a telephone reminder; two, a mailed reminder; and two, no reminder. The percentage of diaries returned in each cell is shown in Table 1. There is considerable variability in compliance, ranging from a high of 93% in the lottery-telephone reminder group to a low of 52 per cent in the no-lottery–mail-reminder group. Apparently, the mailed reminder had no, or a negative, effect on diary returns; by contrast, the telephone reminder increased compliance from 11 to 16 per cent. The lottery ticket had a similar effect on compliance.

Based on the results of this pilot study, both telephone reminder and the lottery ticket were incorporated in the main study.

Compliance Over Two Years—Main Study Results

Six months after the start of the pilot study, the remaining 401 subjects entered the study. Because subjects were followed for 24 months after the initiation of the pilot study, diary information is available for 24

TABLE 1. Pretest Results—Compliance With Different Reminders and Incentives (Diaries Returned/Diaries Sent)

	Lottery Ticket		
	Yes	No	Average
Phone	93% n = 18	74% n = 21	83%
Reminder card	80% n = 18	52% n = 18	67%
None	72% n = 20	72% n = 22	72%
Average	81%	67%	

$\chi^2 = 17.91$ $p < 0.001$.

months on the initial pretest sample of 117 subjects, and for 18 months on the remaining 401 subjects. One further experimental manoeuver, the discontinuation of the telephone reminder, was introduced after 18 months. This decision was taken because there was some indication that subjects were irritated by the repeated telephone reminders. As a result, while the pretest sample was followed for four six-month intervals, they received telephone reminders for the first three intervals but did not in the last six-month interval. The main study sample was followed for 18 months, with telephone reminders during the first 12 months.

At the completion of each six-month period, some individuals refused to be reinterviewed. Six months after the initial interview, 59 (11 per cent) dropped out, and a further 31 individuals (6 per cent) did not complete a third interview. Of the original 117 individuals in the pilot study, 34 (29 per cent) had dropped out by the time of the fourth interview.

Diary compliance was calculated in each six-month sequence for those individuals who consented to be interviewed at the start of the time period. The compliance rates are shown in Table 2. Those rates remained nearly constant over the first 18 months of the study. The deletion of telephone follow-up resulted in a statistically significant drop in compliance of about 5 per cent [$\chi^2 = 40.15$, $p < 0.0001$]. Although significant, the reduced compliance is much less than that which would have been anticipated from the pretest results if telephone reminders had not been implemented at the beginning of the study.

These average values may obscure a wide range in behavior of individual subjects, however. To approach this question, subjects in each time period were categorized into those who had returned fewer than 25 per cent of the diaries, 26–50 per cent, 51–75 per cent, 76–99 per cent, or 100 per cent of the diaries. These data have remained constant over the three time periods, with slightly more than half the subjects returning all diaries and about 80 per cent of subjects returning more than 75 per cent of the diaries. A drop of approximately 2 per cent in the number of subjects returning more than 75 per cent of the diaries was noted when telephone reminders were discontinued.

Further analysis was conducted to determine whether there was any identifiable difference in sociodemographic variables between compliant and noncompliant subjects. The 422 subjects who completed more than 75 per cent of diaries in the first time period were contrasted with the 96 remaining subjects.

Compliant subjects were older (41.4 vs 36.6 years, $t= 3.89$, $p< 0.0001$), and slightly more of the compliant subjects were married [84 per cent vs 76 per cent, $\chi^2 = 7.5$, $p< 0.01$]. Educational level, sex, income and socioeconomic status did not differ between the two groups.

TABLE 2. Diary Compliance Over Four Six-Month Time Intervals

Time Period	Telephone Reminder	Number of Subjects	Number of Diaries Sent	Number of Diaries Received	Compliance
0–6 months	Yes	518	6031	5137	85.2%
7–12 months	Yes	459	6056	5281	87.2%
13–18 months	Yes	92	1192	1024	85.9%
	No	336	4351	3603	82.8%
19–24 months	No	83	1078	878	81.4%

χ^2 phone–no phone = 40.15; $p < 0.0001$.

Dimensions of Health Reported From Diaries

The diary data over the four time periods were analyzed to characterize the dimensions of health revealed by the diaries and to determine the relation between health reporting, compliance and seasonality. Mean values of a number of diary variables in each time period are shown in Table 3. On the average, subjects returned diaries covering from 30 to 34 days within each six-month period. At least one symptom was reported for approximately one third of these days. There is some evidence of a fall off in reporting, amounting to about 12 per cent per year. On those days when symptoms were reported, about 1.5 symptoms were mentioned, with an average severity of 2.6 on a five-point scale.

Although patients reported symptoms on about one day in three, they took action about their symptoms on about two thirds of the days when symptoms were reported. Some activity change was reported for 2.4–2.8 days in each time period, roughly one quarter of those days in which symptoms were reported. Relatively few of the reported symptoms were sufficiently severe to result in time off from work (0.4–0.5 days per period), and few symptoms resulted in a health professional contact (0.5–0.6 days). Overall, patients reported an average health evaluation of 4.8 to 5.4 on a seven-point scale. Not surprisingly, the correlations among these variables are high, ranging from 0.4–0.8.

Although the compliance with diary returns was high, it is of interest to determine whether the results were biased by nonreturns. Two sources of bias could potentially influence the results: 1) individuals who return only a portion of the diaries in any time interval may show different patterns of health from those who return all diaries, and 2) approximately, 5–10 per cent of subjects refused to be reinterviewed after each time period. If these individuals were less healthy than those who were interviewed, this may account for the apparent improvement in the health measures of Table 3 over successive six-month intervals.

The effect of nonreturns within a time period was addressed using the two groups designated above. Those individuals who

TABLE 3. Mean Values of Diary Variables over Four Six-Month Time Sequences

	Sequence I	Sequence II	Sequence III	Sequence IV	Significance
Sample size	518	459	428	83	
Average number of diaries sent	34.93	39.58	38.85	38.96	
Average number of diaries	29.96	34.42	34.19	34.51	<0.0001
Number of days with symptoms	10.8	10.5	9.3	9.1	0.06
Per cent of days with symptoms	0.36	0.28	0.27	0.26	<0.0001
Number of symptoms per symptom day	1.46	1.40	1.47	1.51	NS
Average severity (5 = point scale)	2.67	2.69	2.66	2.65	NS
Number of days with action	6.99	6.58	5.89	6.04	NS
Number of days with activity change	2.51	2.46	2.40	2.83	NS
Number of days with health professional contact	0.62	0.66	0.59	0.58	NS
Number of days off from work	0.58	0.49	0.56	0.42	NS
Average health evaluation (7 = point scale)	4.81	5.45	5.44	5.37	<0.0001

returned more than 75 per cent of diaries were designated the "compliant" group, and individuals returning 75 per cent of fewer diaries were designated a "noncompliant" group. The compliant group returned an average of 33.8 diaries in the first sequence, 96 per cent of those sent. By contrast, the noncompliant group returned 11.9 diaries on the average, a return rate of 36 per cent. The noncompliant subjects naturally reported fewer symptoms and actions overall; the proportion of reported days with symptoms, activity change, time off from work and health professional contact, however, did not differ between the two groups. Thus, compliance with diary returns does not appear to influence subjective perception of health, and the information received from relatively noncompliant subjects may be a valid sample of their perceived health.

To determine if dropouts from the study resulted in a bias, the analysis of Table 3 was repeated using only the 428 subjects who remained in the study over all three sequences. The data were similar to those of Table 3—in particular, a similar trend to lower reporting of health-related behavior over time was noted. This trend is therefore likely due to the fall-off effect described by Mooney,[4] rather than to a selective dropout of less healthy individuals from the sample.

The data on reported symptoms and actions were also analyzed by calendar month over the two-year duration of the study. The proportion of reported diary days with symptoms and action showed similar trends over time. A seasonal effect was present, with a sharp minimum in June, July and August of each year, during which approximately 20–25 per cent of days had reported symptoms and about 15 per cent resulted in some action. These variables rose to a maximum of about 45 per cent symptom days and 25 per cent of days with action in the fall months, followed by a gradual decline over the winter and spring months.

Relationship between Diary and Other Health Measures

To examine the relationship between measures of health obtained from the diary and other measures of health and illness used in the study, selected diary variables were correlated with 1) the number of symptoms obtained from the Langner scale administered at each interview and 2) number of visits to a family physician. The Langner scale of subjective distress has frequently been used in the literature as a measure of the effect of stress on individuals and was found in the present study to be very stable over time (r= 0.85 over successive time sequences).

Most of the diary variables show a moderate correlation with the Langner score, ranging from 0.20 to 0.35. The number of days off from work and the number of days with a health professional contact exhibit much lower correlations; however, the distribution of these two variables is highly skewed, with 68–80 per cent of patients reporting no events in these categories.

The other measure of health available is the number of encounters with the patient's family physician. In contrast to health professional contacts reported in the diary, these data were obtained directly from the subject's family physician using a brief encounter form designed for the study. The number of encounters was highly skewed, with 55 per cent of patients reporting no encounters in a six-month period. As a result, the correlations between the diary variables and physician utilization ranged between 0.14 and 0.26, with the best correlation being between the reported number of diary days with health professional contact and the number of family physician visits. Because the reported health professional contacts from the diary represent only a sample of family physician visits (on those days sampled by the diary) and cover visits to all health professionals, this low correlation is not too surprising.

Conclusions

The health diaries used in the present study have served a useful role in measuring perceived morbidity in a large sample over an extended time period. The use of several strategies—time sampling at random intervals, lottery tickets and telephone reminders—maintained high compliance over the duration of the study and reduced the conditioning effect to a reasonable level, resulting in fewer reported symptoms.

The data obtained from the diaries result in a dramatically different perspective on health and illness than that obtained from physician utilization data. Although subjects visited a family physician only once, on the average, in a six-month interval, and half the sample reported no visits in each time period, they reported symptoms on about one day in three. Apparently, most of the reported symptoms are minor, since about one third of the days with symptoms resulted in no action being taken by the subject. Some activity change was reported for about 25 per cent of the symptom days, but less than one day of the 30 days sampled in each time period resulted in time off from work or a physician visit. Variables derived from the diary reports showed a moderate correlation with a symptom scale administered at the time of the home interview and a low correlation with visits to the family physician.

Apparently, the health diaries represent a sensitive measure of a dimension of illness that could not be assessed by other means. In terms of the major objectives of the study, it remains to be shown whether aspects of the psychosocial environment, life stress, and social personality variables are more directly related to the frequent occurrence of minor illness documented by the health diary, or to the more severe episodes of illness that are brought to the family physician.

References

1. White KL, Williams TF, Greenberg BG. The ecology of medical care. N Engl J Med 1961;265:885.
2. Allen GI, Breslow L, Weissman A, et al. Interviewing versus diary keeping in eliciting information in a morbidity survey. Am J Public Health 1954; 44:919.
3. Verbrugge LM. Health diaries. Med Care 1980;18:73.
4. Mooney HW. Methodology in two California health surveys. Public Health Monograph No. 70. P.H.S. Publication No. 942. Washington, D.C., 1962.
5. Rahe RH. The pathway between subject's recent life changes and their near-future illness reports. In: Dohrenwend BS, Dohrenwend BP, eds. Stressful life events: their nature and effects. New York: Wiley, 1974:73–86.
6. Rotter JB. Social learning and clinical psychology. Englewood Cliffs, N.J.: Prentice Hall, 1954.
7. Wallston KA, Maides S, Wallston BS. Health related information seeking as a function of health-related locus of control and health value. Journal of Research in Personality 1976;10:215.
8. Rokeach M. The nature of human values. New York: Free Press, 1973.
9. McFarlane AH, Norman GR, Streiner DL, et al. Methodological issues in developing a scale to measure social support and some findings. Schiz Bull 1981;7:90.
10. Langner TS. A twenty-two item screening score of psychiatric symptoms indicating impairment. J Health Soc Behav 1962;3:269.
11. McFarlane AH, Norman GR, Streiner DL, et al. A longitudinal study of the influence of the psychosocial environment on health status: the multiple dimensions of life changes. J Health Soc Behav 1980;21:124.

MEDICAL CARE
June 1982, Vol. XX, No. 6

Patient Characteristics That Influence Physician Behavior

ELIZABETH M. HOOPER, B.S.N., LORETTO M. COMSTOCK, M.S., JEAN M. GOODWIN, M.D., M.P.H., AND JAMES S. GOODWIN, M.D.

We examined whether four patient characteristics (age, ethnicity, sex and appearance) influenced five physician behaviors with the patient (interviewing, nonverbal attention, courtesy, information giving and empathy). Ten physician–patient interactions were observed through a one-way mirror for each of 15 physicians. Physicians were rated higher on information giving ($p = 0.002$) and empathic skills ($p = 0.02$) when they were with female patients; on interviewing ($p = 0.02$) and empathy ($p = 0.0007$) with Anglo-American compared with Spanish-American patients; on interviewing ($p = 0.01$), nonverbal attention ($p = 0.001$) and courtesy ($p = 0.02$) when with the best groomed patients; and on courtesy ($p = 0.03$) with elderly patients. In addition, there were fewer physician-initiated interruptions with female patients ($p = 0.03$) and with well-groomed patients ($p = 0.02$). Each of the four patient characteristics studied acted independently in influencing physician behavior. The recognition that patient characteristics influence physician behavior should stimulate physicians to examine their reactions in order to insure that all types of patients receive thorough, courteous and empathic care.

LITTLE HAS BEEN DOCUMENTED about how a patient's attributes or characteristics directly affect a physician's behavior. This paucity of research reflects a medical ideal that all patients should be treated equally. "The physician should see to it that the patient's sentiments do not act upon his sentiments and, above all, do not thereby modify his behavior . . ."[1] Despite twentieth-century research showing that physician–patient interactions function according to the same rules as do other social systems,[2–4] the ideal persists. In 1968, when 132 physicians were asked about their behavior with patients, 80 per cent stated that they "gave the same amount of attention to all patients."[5] More accurate and less idealistic perceptions of how physicians react to patient characteristics may allow the physician to improve patient compliance and to identify patients' problems more realistically.

Recently there has been increasing interest in several areas of patient–physician interaction.[6,7] Eisenberg points out that most of the work in the field suffers from lack of case-controlled data and from an excess of opinion over fact.[8] Nevertheless, some investigations have been well

---

From the Departments of Medicine and Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico.

This paper has been presented in part at the Western Section meetings of the American Federation for Clinical Research, Carmel, California in February 1980 and again at the National Meeting of the American Federation for Clinical Research, May 1980, in Washington , D.C.

Address for reprints: James S. Goodwin, M.D., Department of Medicine, University of New Mexico School of Medicine, Albuquerque, NM 87131.

0025-7079/82/0505/0630/$00.95 © J. B. Lippincott Co.

controlled. The few studies that have focused on the effects of patient characteristics on medical care have used such indirect measures of physician behavior as number of medications, referral to specialists, number of operations performed, or whether a physician's assistant or a physician is the primary caretaker. For example, Penchansky and Fox reported that primary physicians in group practice referred white patients to consultants more frequently than black patients.[9] Egbert and Rothman found that blacks were two to four times as likely as whites to have surgery performed by a resident rather than a staff surgeon, and this difference remained when diagnosis and mode of payment were controlled for.[10] Medicaid recipients were also at higher risk for receiving care by physicians in training. Perkoff and Anderson studied ward versus private patient status for patients visiting an emergency room and found that black patients with similar medical insurance were more likely than whites to be admitted to ward status and that whites were more likely to be admitted to the hospital than blacks with similar chief complaints.[11] Other studies suggest that socioeconomic status affects how a patient is treated. The higher a patient's socioeconomic status, the more likely he/she is to be seen by a physician rather than paramedical personnel,[12] the more consultations are made, and the more likely he/she is to receive an appendectomy or tonsillectomy.[13]

In addition to race and socioeconomic status, some authors have postulated that certain types of patients might elicit negligent behavior from physicians. For example, articles by Lennare and Lennare[14] and Howell[15] allege that women are treated differently than men on the basis of sexual prejudice. Papper simply states that aged, dirty and ungrateful patients are undesirable to treat and may receive inferior care.[16] Duff and Hollingshead concluded that a patient's physical appearance influences medical care[4] and Baldigo et al. have shown that medical students attribute negative characteristics to obese patients.[17] Crane emphasizes the importance of the physician's individual reaction to the patient's disease in determining the course of treatment.[18]

Our study examines how the characteristics of the patient affect the behavior of the physician. We did not look at end points or outcome; we looked directly at physician behaviors during the patient–physician interaction. The patient attributes we chose to study were age, ethnicity, sex and neatness of appearance. These are all attributes routinely noted in the first sentence of the medical history. The physician behavoirs assessed were information giving, interviewing, empathy, nonverbal attention and courtesy. We found marked differences in these physician behaviors depending on characteristics of the patient.

Method

Subjects

This study was conducted in the outpatient clinic of a university teaching hospital. Fifteen unselected residents in their second and third year of medical training were studied in their weekly outpatient clinics. All physicians approached agreed to participate in the study. One hundred fifty physician–patient interactions (ten per physician) were evaluated through a one-way mirror by a trained observer over a three-month period. We approached 162 patients, and 150 agreed to participate. Written informed consent was obtained from both the patient and the physician. All 15 physicians were of Anglo-American (non-Hispanic white) ethnic origin; four were female. One hundred seven of the patients were female, and 43 were male. Sixty-seven patients were of Anglo-American origin, 74 were Spanish-American, three were American Indian and six were black. All of the study population spoke English. Mean patient age was 57.5 years; all physicians were between 25 and 35 years of age.

Observations

Each of 28 physician behaviors was rated on a four-point scale and subsequently grouped for analysis into five categories as described in the results section of this article.[19] Physician behaviors rated included such skills as listening, information giving, use of humor, empathy, etc. If a skill was not observed, no rating was recorded for that skill, and it was treated statistically as a missing value. For example, use of humor by the physician and/or patient was rated for its appropriateness and effectiveness. Use of light humor that helped put the patient at ease would receive a favorable rating, whereas a physician's joking response to a patient's expression of fear toward an upcoming procedure would receive an unfavorable rating. If neither the patient nor the physician initiated any humor, no score was recorded. Observer variability was assessed by having two observers rate the same interactions independently and comparing the ratings. In 122 of 130 ratings (94 per cent), the two observers did not differ or differed by only one point in scoring the physician behavior. Data collected in addition to the 28 behaviors include patient age, total patient time with physician, number of physician-initiated interruptions (when the physician left the room), and sex and ethnicity for both the patient and physician. Physical appearance was also rated on a four-point scale. A rating of 4 would describe someone who was "clean and pressed" in a three-piece suit or attractive dress, hair clean and neatly styled. A rating of 1 would describe an individual with rumpled and dirty clothing who was generally disheveled.

Statistical Analysis

All variables were keypunched directly onto IBM cards from the observation forms. Data were analyzed using Statistical Analysis System (SAS) version 79.2B, and Biomedical Data Processing (BMDP) on an IBM 3032 computer. Inferential analysis was done using Student's t-test for comparing two groups, one-way analysis of variance for comparing more than two groups, and a log-linear model for analysis of variance of categorical data to test for interrelationships among variables.

Results

The ratings of 28 physician behaviors were grouped into five categories. "Information giving" assessed the effectiveness of the physician in communicating verbally, clarifying, summarizing and talking at the patient's level. "Interviewing" included such skills as using open-ended questions, eliciting details and allowing the patient an opportunity to ask questions. "Empathy" addressed the physician's ability to acknowledge the patient's prominent feelings, to pick up on the patient's mood changes (for example, if a patient's behavior were to change from loquacious to quiet with closed body language), and to acknowledge the patient's emotional needs. "Nonverbal attention" included eye contact and the physician's position in relation to the patient. "Courtesy" assessed the physician's formal greeting of the patient and formal closure of the interaction.

The mean ratings for the observed physician behaviors are listed in Table 1. Nonverbal attention to the patient, as well as interviewing skills, were rated uniformly high. Skills with intermediate ratings included information giving and courtesy. With regard to courtesy, greetings were rated higher than closures (3.2 ± 0.05 vs 2.8 ± 0.08, mean ± SEM, $p < 0.01$).

In general, the ratings for empathic skills, such as picking up on the patient's mood changes and acknowledging the patient's feelings and emotional needs, were lowest. Two other areas with low scores were the use of humor and touching.

Humor was often used inappropriately and was seen primarily as a release from tension when either the patient or physician was uncomfortable. Some physicians touched patients as a comforting gesture, but not all patients were comfortable receiving this touching. Some patients appeared eager for a comforting gesture from the physician, but not all physicians perceived this cue or were able to respond to it.

Physician Behavior and Patient's Sex

As shown in Table 2, ratings of information giving and empathy were significantly higher with female patients than with male patients, while there were no differences in the ratings of interviewing, nonverbal attention or courtesy toward male and female patients. In addition, physician-initiated interruptions were significantly less frequent with female patients.

Physician Behavior and Patient Ethnicity

As shown in Table 3, the ratings of interviewing and empathy behaviors were significantly higher with Anglo-American than with Spanish-American patients. In addition, there was a trend for the physician to spend more time with the Anglo-American than with the Spanish-American patient (p = 0.07). Thus, ratings of physician skills that involve receiving and perceiving communications by the patients (interviewing and empathy) were significantly lower with the Spanish-American patients, while those of communicative skills not requiring the sensitive processing of the patient's input (nonverbal attention, courtesy, information giving) were not significantly different. Our feeling was that the Anglo-American physician and Spanish-American patients attempted to communicate with each other, but the communication was not fully received by either party. It should be emphasized that all the Spanish-American patients spoke English; however, for most of these patients Spanish was the primary language used in the family.

TABLE 1. Mean Rating of Physician Behaviors by the Observer

Observations	Rating *
Interviewing	3.5 ± 0.7
Attention	3.4 ± 0.7
Courtesy	3.2 ± 0.7
Information giving	3.0 ± 1.0
Empathy	2.0 ± 1.1
Physician-initiated interruptions	0.5 ± 0.8
Time of interaction	16.4 ± 8.6
Patient appearance	3.2 ± 0.9

* Data given as mean ± SD for 150 physician–patient interactions.

Physician Behavior and Patient Appearance

The two lowest scores for patient appearance were combined into one for analysis. As shown in Fig. 1, all ratings improved as the patient's appearance improved. For three ratings (interviewing, nonverbal attention and courtesy) this difference was statistically significant. There were also fewer physician-initiated interruptions with the patients rated highest on appearance (p = 0.02). There was no difference in total time of the physician–patient interaction as a function of patient appearance (p > 0.5).

TABLE 2. Physician Behaviors With Male and Female Patients

Observations	Female Patients (106)	Male Patients (42)
Interviewing	3.5 ± 0.6*	3.4 ± 0.7
Attention	3.4 ± 0.7	3.3 ± 0.7
Courtesy	3.2 ± 0.7	3.3 ± 0.7
Information giving	3.1 ± 0.9†	2.5 ± 1.0
Empathy	2.2 ± 1.1†	1.7 ± 0.9
Physician-initiated interruptions	0.4 ± 0.8‡	0.7 ± 0.8
Time of interaction	16.4 ± 9.1	17.3 ± 7.8

* Data given as mean ± SD.
† Different from male patients, p = 0.002.
‡ Different from male patients, p < 0.03.

TABLE 3. Physician Behavior With Anglo-American and Spanish-American Patients

Observations	Anglo-American Patients (67)	Spanish-American Patients (74)
Interviewing	3.6 ± 0.6*†	3.3 ± 0.7
Attention	3.4 ± 0.7	3.3 ± 0.7
Courtesy	3.2 ± 0.7	3.1 ± 0.7
Information giving	3.0 ± 0.9	2.8 ± 1.0
Empathy	2.4 ± 1.0‡	1.7 ± 1.1
Physician-initiated interruptions	0.6 ± 0.9	0.4 ± 0.7
Time of interaction	18.0 ± 10.2	15.3 ± 7.1

* Data given as mean ± SD.
† Different from Spanish-American patients, $p = 0.02$.
‡ Different from Spanish-American patients, $p < 0.001$.

Physician Behavior and Patient Age

Patients were divided into four age groups, under 45, 45–64, 65–74 and over 74 years of age. As shown in Fig. 2, there were no changes in ratings of physician behaviors with the age of the patients except for those patients over age 74, when all

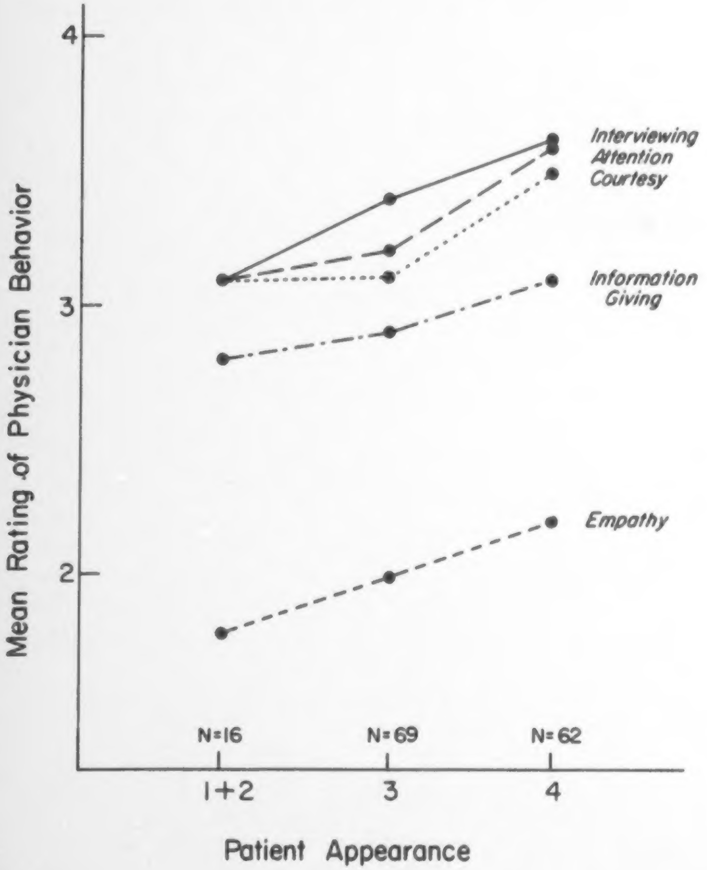


FIG. 1. Physician behaviors as a function of patient appearance. As patient appearance improved, physicians were rated significantly higher at interviewing ($p < 0.01$), attention ($p < 0.001$) and courtesy ($p < 0.02$) skills. The ratings for information giving and empathy did not vary significantly with patient appearance ($p > 0.4$). Data were analyzed by one-way analysis of variance.

ratings of physician behaviors tended to improve. The increased ratings with these elderly patients were significant only for courtesy (p = 0.03).

When the patient population was divided at age 40, there were two significant differences. More time was spent (p = 0.003), and ratings of information giving increased (p = 0.01) with the older population.

Independence of Patient Characteristics in Influencing Physician Behavior

Because we found significant differences in our ratings of some physician behaviors as a function of the patients' sex, ethnicity, appearance and age, we tested for interrelationships among these patient characteristics. For example, we asked whether ratings of physician courtesy, which were significantly higher for patients over the age of 75, could have been due to an increased number of female, Anglo-American and/or neat appearing individuals in this age group. Using a log-linear model for analysis of variance of categorical data, we found that there were no significant interrelationships (p > 0.40 for each of the 90 possible interrelationships). This suggests that each of the patient characteristics that we measured acted independently in influencing the ratings of physician behavior.

Discussion

The results of this study are that ratings of physician behaviors, made by a trained observer, varied with the sex, age, appearance and ethnicity of the patient. Physicians were rated significantly higher on in-

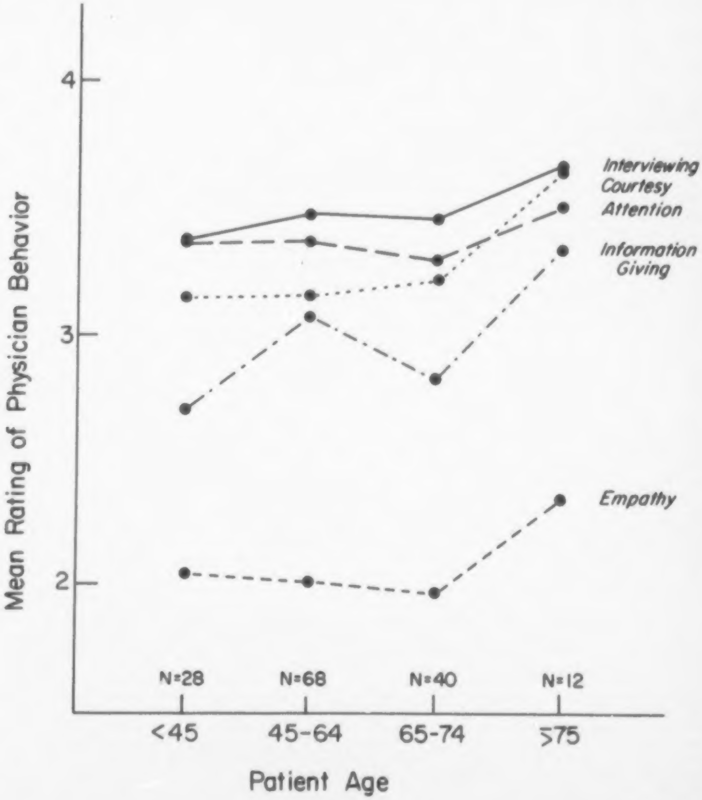


FIG. 2. Physician behavior as a function of patient age. Ratings of all behaviors tended to increase in patients over the age of 75. This difference was significant only for the ratings of courtesy (p = 0.03).

formation giving and empathy when they were dealing with female patients; on interviewing and empathy with Anglo-American patients; on interviewing, attention and courtesy when with the best groomed patients; and on courtesy with elderly patients. Although we did show a high degree of interobserver reliability, we have not ruled out the possibility of observer bias due to shared perceptions. For example, the two observers may share the perception that older patients are treated "better" or that Spanish-American patients are discriminated against. The fact that the objective data collected (time of interview and number of physician-initiated interruptions) supported the subjective data would tend to strengthen our overall findings.

The four patient characteristics all acted independently in influencing physician behavior. Each characteristic influenced physician behavior to the same extent before and after we controlled for the effects of the other patient characteristics. On the whole, then, well-groomed, elderly Anglo-American women received the highest rated physician behaviors, while disheveled, young Spanish-American males received the lowest rated behaviors. It is still possible, however, that the various influences on physician behavior by the four patient characteristics were mediated in part by other characteristics that were not examined in this study as, for example, the patient's diagnosis or educational status.

Our findings on changes in ratings of physician behavior with the sex of the patient appear to disagree with the widely publicized view that medical care is given in a way that disparages women.[15,20] We found that physicians were rated significantly higher in information giving and empathy with female than with male patients. In addition, physicians left the room during the interview significantly fewer times with female patients. These data are in accord with those of Keskiner et al. who found that female patients were more likely to be referred to a psychiatrist than were male patients with the same symptoms of neurosis;[21] in both studies possible sexist attitudes of the physicians would appear to work in favor of better medical care for female patients.

The decreased rating of physician interviewing and empathy skills with Spanish-American patients was expected and, to some extent, unavoidable. Intimate communication across a cultural barrier is more difficult than within the same culture, even when both individuals speak the same language.[22,23]

The generally low rating for empathetic skills with all patients supports Duffy's data showing that one of the least frequently used skills by physicians is the ability to ascertain patient emotions and that physicians are least comfortable speaking about psychosocial problems.[24]

It should be emphasized that the fact that we found differences in ratings of physician behavior with the different types of patients does not necessarily prove that these differences represent inappropriate behavior by the physicians studied. The purpose, after all, of the patient's visit to the physician is the provision of medical care, not an exercise in courtesy, communication or empathy. In addition, the changes in ratings of physician behavior could result from the physician's perceptions of the patient's expectations. For example, the increased rating of courtesy toward elderly patients could be seen as an appropriate response to the expectations of this group of patients. On the other hand, data from other studies and our own preliminary data show that such physician behaviors as courtesy, interviewing and information giving do correlate with patient satisfaction[25–30] and with subsequent patient compliance with their medical regimens.[26–28] Therefore, it is legitimate to ask whether the decrease in ratings of interviewing skills when physicians were working with Spanish-American patients, for example, might contribute to poorer medical care for these individuals.

The concept that a physician's behavior is influenced by a particular characteristic of the patient has evolved in the past two decades from an initial position of outright denial. Until recently, medical students and physicians were simply told that it was wrong to let a patient's personal characteristics influence their behaviors.[1,31] More recently it has become acceptable to recognize that extremes of patient characteristics and behavior do influence the physician. Thus, there has been a series of articles advising physicians on how to deal with "problem patients," "disliked patients" or "hateful patients."[32-35] We have recently reported that patients with impaired mental status are disliked by their physicians, though the physicians were not consciously aware of this association.[35] In addition, we have shown that particularly stressful patients are at risk for punitive procedures such as inappropriate placebo use.[36]

Many physicians are now willing to admit that certain types of patients (for example, alcoholics) give them trouble. We have attempted to take this one step further, to show that different patient characteristics, none of them "hateful," have a measurable influence on ratings of physician behavior. This observation should be obvious to sociologists, who have long recognized that all social interactions are bi-directional.[2-4] In addition, psychiatrists since Freud have worked within the constructs of transference and countertransference, recognizing that particular characteristics of a patient can trigger unconscious conflicts in the physician and adversely affect the patient's psychotherapy.[37] Obvious though this may seem, medicine has managed to ignore, almost entirely, the fact that patient characteristics influence physician behavior. Recognition by physicians that they behave differently with different types of patients can stimulate them to examine their reactions in an attempt to insure that all patients, regardless of age, sex, race and appearance, receive thorough, courteous and empathic medical care.

Acknowledgments

The authors thank Jonathan Abrams, Robert Kellner, Jonathan Samet, Albert Vogel and Ralph Williams for their suggestions on improving the manuscript; Kathleen Aragon for her editorial assistance and Dorothy Pathak for her assistance in the statistical tests on the data.

References

1. Henderson LJ. Physicians and patient as a social system. N Engl J Med 1935;212:819.

2. Bloom SW. The doctor and his patient: a sociological interpretation. New York: Free Press, 1965.

3. Freidson E. Profession of medicine: a study of the sociology of applied knowledge. New York: Dodd, Mead & Co., 1973.

4. Duff RS, Hollingshead AB. Sickness and society. New York: Harper & Row, 1968.

5. Davis MS. Attitudinal and behavioral aspects of the doctor–patient relationship as expressed and exhibited by medical students and their mentors. J Med Educ 1968;43:337.

6. Brody DS. The patient's role in clinical decision making. Ann Intern Med 1980;93:718.

7. Freiden RB, Goldman L, Cecil RR. Patient–physician concordance in problem identification in the primary care setting. Ann Intern Med 1980; 93:490.

8. Eisenberg JM. Sociologic influences on decision making by clinicians. Ann Intern Med 1979;90:957.

9. Penchansky R, Fox D. Frequency of referral and patient characteristics in group practice. Med Care 1980;8:368.

10. Egbert LD, Rothman IL. Relation between the race and economic status of patients and who performs their surgery. N Engl J Med 1977;297:90.

11. Perkoff GT, Anderson M. Relationship between demographic characteristics, patient's chief complaint, and medical care destination in an emergency room. Med Care 1970;8:309.

12. Rushing WA, Miles DL. Physicians, physicians' assistants and the social characteristics of patients in southern Appalachia. Med Care 1977; 15:1004.

13. Allander E. Distribution of doctors' consultations, operations, and symptoms with special relevance to social factors. Scand J Soc Med 1974;2:37.

14. Lennare KJ, Lennare RJ. Alleged psychogenic disorders in women—a possible manifestation of sexual prejudice. N Engl J Med 1973;288:288.

15. Howell MD. What medical schools teach about women. N Engl J Med 1974;291:304.

16. Papper S. The undesirable patient. J Chronic Dis 1970;22:777.

17. Baldigo J, Breytspraak L, Cohen-Conger J, et al. Impression formation in the patient interview: a training experience with first year medical students. Proceedings of the Fourteenth Annual Conference on Research in Medical Education, Washington, D.C., Association of American Medical Colleges, 1975:206.

18. Crane D. The sanctity of social life: physicians' treatment of critically ill patients. New York: Russell Sage Foundation, 1975.

19. Comstock LM, Williams RC. The way we teach students to care for patients. Medical Teacher 1980;2:168.

20. Schiefelbein S. The female patient heeded? hustled? healed? Saturday Review, March 29, 1980.

21. Keskiner A, Zalcman MJ, Ruppert EH. Advantages of being female in psychiatric rehabilitation. Arch Gen Psychiatry 1973;28:689.

22. Levine CC. Doctor–patient communication with the inner-city adolescent. N Engl Med 1970;282:454.

23. Kleinman A, Eisenberg L, Good B. Culture, illness and care: clinical lessons for anthropologic and cross-cultural research. Ann Intern Med 1978;88:251.

24. Duffy DL, Hammerman D, Cohen MA. Communication skills of house officers. Ann Intern Med 1980;93:354.

25. Comstock LM, Hooper E, Goodwin JM, Goodwin JS. Physician behaviors that correlate with patient satisfaction. Clin Res 1980;28:293A.

26. Freeman B, Negrete VF, Davis M, Korsche BM. Gaps in doctor–patient communication: doctor–patient interaction analysis. Pediatr Res 1971;5:298.

27. Korsch BM, Gozzi EK, Francis V. Gaps in doctor–patient communication. Pediatrics 1968; 42:855.

28. Davis MS. Variation in patients' compliance with doctors' orders: medical practice and doctor–patient interaction. Psychiat Med 1971;2:31.

29. Stiles WB, Putnam SM, Wolf MH, James SA. Interaction exchange structure and patient satisfaction with medical interviews. Med Care 1979;17:667.

30. Bolton B, Sommer P. Mode of address and patient satisfaction in rehabilitation: an experimental study. J Health Soc Behav 1970;11:215.

31. Daniels MJ. Affect and its control in the medical intern. Am J Soc 1960;16:259.

32. Strain JJ, Hammerman D. Ombudsmen (medical-psychiatric) rounds. Ann Intern Med 1978;88:550.

33. Drossman DA. The problem patient. Ann Intern Med 1978;88:366.

34. Groves JE. Taking care of the hateful patient. N Engl J Med 1978;298:883.

35. Goodwin JM, Goodwin JS, Kellner R. Psychiatric symptoms in disliked medical patients. JAMA 1979;241:1117.

36. Goodwin JS, Goodwin JM, Vogel AV. Knowledge and use of placebos by house officers and nurses. Ann Intern Med 1979;91:106.

37. Freud S. The future prospects of psychoanalytic therapy. In: Collected Papers, Vol. 2, New York: Basic Books, 1959;285.

MEDICAL CARE
June 1982, Vol. XX, No. 6

Communications

Do Postcard Reminders Improve Influenza Vaccination Compliance?

A Prospective Trial of Different Postcard "Cues"

ERIC B. LARSON, M.D., M.P.H., JAMES BERGMAN, M.D., FRED HEIDRICH, M.D.,
BARBARA L. ALVIN, M.S., AND RONALD SCHNEEWEISS, M.D.

A randomized trial of various postcard reminder "cues" was performed to improve understanding of health-related behavior and to find better strategies for improving influenza vaccination compliance. Data were gathered on 283 high-risk patients (92 per cent response rate) who received: 1) a "neutral" card simply announcing the availability of vaccine; 2) a "Health-Belief-Model" card written to take advantage of the association between certain health beliefs and vaccination behavior; 3) a "personal" card signed by the patient's physician; or 4) no postcard. The highest rate of vaccination occurred among recipients of the Health-Belief-Model postcard (51.5 per cent vs. 20.2 per cent for control, $p < 0.001$). Linear logistic regression analysis found that age, prior vaccination history and experimental group had a significant effect on likelihood of being vaccinated. After adjusting for age and prior vaccination experience, the vaccination rate was found to be significantly higher for persons receiving the Health-Belief-Model postcard compared with persons receiving no postcard or a neutral postcard. We conclude that reminder postcards emphasizing elements of the health belief model may help increase vaccination rates.

ANNUAL INFLUENZA IMMUNIZATION has been recommended since 1964 for persons over the age of 65 and for persons with chronic cardiac, pulmonary, renal and metabolic diseases. Immunization is intended to prevent the excess mortality associated with influenza epidemics in this high-risk group.[1] However, overall vaccination rates among high-risk patients have been as low as 10 to 15 per cent according to the Center for Disease Control (CDC).[2]

These low rates have prompted us to study influenza vaccination to learn more about the general subject of health behavior and to find effective means to promote healthy behavior.[3] Although research efforts over the years have used different constructs to explain the effects of personal, psychosocial and environmental elements on health and illness behavior,[4–8] we chose to study the Health Belief Model (HBM) as formulated by Hochbaum, Becker, Rosenstock and others.[8]

From the Departments of Medicine, Community Medicine, Family Medicine and Biostatistics, University of Washington, Seattle, Washington.

This research was presented at the American Federation Clinical Research National Meeting, May 12, 1980.

This work was supported in part with BSRG research funds from the University of Washington School of Medicine. Dr. Larson was George Morris Piersol Teaching and Research Scholar of the American College of Physicians and is a Henry J. Kaiser Family Foundation Faculty Fellow in General Internal Medicine.

Reprint requests: Eric B. Larson, M.D., M.P.H., University of Washington School of Medicine, Department of Medicine RG-20, Seattle, WA 98195.

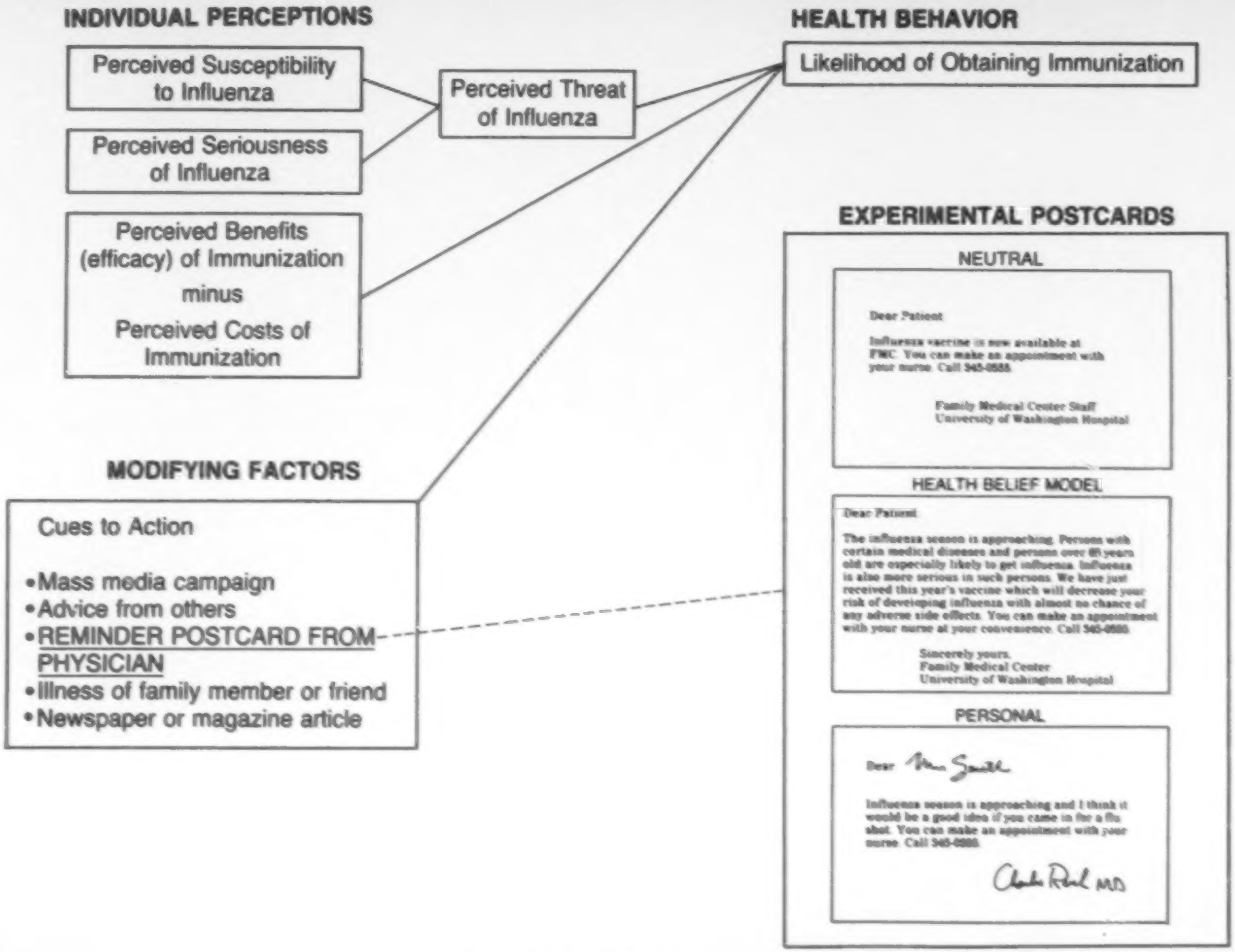
0025-7079/82/0505/0639/$01.00 © J. B. Lippincott Co.

The HBM is a blend of social and psychologic theories, particularly value expectancy theory. The model predicts that an individual's perception of 1) susceptibility to a disease, 2) severity of the disease, and 3) benefits and costs of the action to prevent the disease will determine the likelihood of the individual undertaking that action. The HBM suggests two common determinants of impulses to action: 1) the value placed by an individual on a particular outcome or goal and 2) the individual's estimate of the likelihood that a particular action will produce the desired outcome.[9]

The model has been tested in a variety of settings involving symptomatic and asymptomatic health behaviors.[10–14] Our initial study of influenza vaccination behavior during the 1975 vaccination period validated a variation of the HBM (Fig. 1), and showed positive correlations with compliance and the elements of the model.[3] The study also demonstrated that a reminder postcard or "cue" was an effective means of promoting vaccination. Our study, like many others, was retrospective, but even in this retrospective study, the model explained a relatively small amount of variance.

Two more recent studies reported higher rates of correlation of the HBM with compliance in the Swine Flue immunization program. Using logit analysis of data collection retrospectively, Rundall and Wheeler[15] reported that the HBM accounted for 34 per cent of the variance in the use of vaccine by senior citizens. Cummings et al.,[16] in a prospective study of Swine Flu vaccination, explained more than 40 per cent of the variance with predictor variables, which included the HBM, measures of behavioral intention, social influence, physician's advice, socioeconomic status and past experience with flu shots. Path analysis revealed that most of the influence of HBM variables on behavior was mediated through behavioral intention and that physician recommendations showed a substantial direct effect on vaccination behavior. Sackett[17] has reported that, in contrast to retrospective studies, prospective studies of the HBM have shown inconsistent results. He believes these results to suggest that patients' health beliefs may result from, rather than cause, compliance. In the Canadian study of compliance with antihypertensive theory, beliefs changed to coincide with compliance: after six months of therapy, but not before, compliant patients perceived hypertension to be a more serious disease that benefited from drug therapy.[18] The findings of Sackett[17,18] and Cummings et al.,[16] therefore, suggest that behavioral intent and subsequent behavior could be influenced by efforts directed at changing relevant health beliefs and support the rationale for our prospective trial of three postcard cues, each containing a different message (Fig. 1).

Postcard cues have been shown to lower broken appointment rates[19–21] and have been retrospectively correlated with vaccination compliance.[3] We reasoned that a postcard containing information emphasizing those health beliefs that correlated with vaccinations might be used to change behavioral intentions, thereby promoting vaccination compliance. We also wondered whether a postcard with a personal message from a patient's primary care physician might improve compliance because of the previously demonstrated effect of physician recommendations on health and vaccination behavior. Accordingly, we designed and executed a randomized prospective trial of the effect of cue postcards on vaccination behavior that compares vaccination rates of patients receiving a "neutral" postcard, an "HBM" postcard, a "personal" postcard, and no postcard. The purpose of the trial was to compare the effect of the HBM postcard with the effect of no postcard or a neutral postcard and to make similar comparisons


INDIVIDUAL PERCEPTIONS
Perceived Susceptibility to Influenza
Perceived Seriousness of Influenza
Perceived Benefits (efficacy) of Immunization
minus
Perceived Costs of Immunization
Perceived Threat of Influenza
HEALTH BEHAVIOR
Likelihood of Obtaining Immunization
MODIFYING FACTORS
Cues to Action
•Mass media campaign
•Advice from others
•REMINDER POSTCARD FROM PHYSICIAN
•Illness of family member or friend
•Newspaper or magazine article
EXPERIMENTAL POSTCARDS
NEUTRAL
HEALTH BELIEF MODEL
PERSONAL

Dear Patient

Influenza vaccine is now available at FMC. You can make an appointment with your nurse. Call 345-0888.

Family Medical Center Staff
University of Washington Hospital

Dear Patient

The influenza season is approaching. Persons with certain medical diseases and persons over 65 years old are especially likely to get influenza. Influenza is also more serious in such persons. We have just received this year's vaccine which will decrease your risk of developing influenza with almost no chance of any adverse side effects. You can make an appointment with your nurse at your convenience. Call 345-0888.

Sincerely yours,
Family Medical Center
University of Washington Hospital

Dear Mr Smith

Influenza season is approaching and I think it would be a good idea if you came in for a flu shot. You can make an appointment with your nurse. Call 345-0888.

MD

FIG. 1. "Health Belief Model" and experimental postcards.

for the personal postcard. The study also allows us to observe trends in influenza vaccination among outpatients following the national "Swine Flu" vaccination debacle of 1976, which we anticipated would lower overall vaccination rate.

Methods

The study was conducted at the University of Washington Family Medical Center (FMC). The study population consisted of those patients identified as being at high risk for serious complications from influenza infection[1]: patients over 65 years of age and/or patients with chronic heart disease, bronchopulmonary disease, renal disease and diabetes mellitus, based on ICDA-8 diagnosis codes stored in the FMC's computer.

Three hundred ninety-five patients were identified and selected for study. These patients were randomly assigned to one of four groups: a control group receiving no postcard, a group receiving a neutral postcard, a group receiving an HBM postcard and a group receiving a personal postcard. When the 1978–1979 influenza vaccine became available, the postcards shown in Fig. 1 were mailed to patients. The HBM postcard obviously emphasizes the severity of influenza, the susceptibility of older persons to influenza and the benefits of vaccination. We did not intend to imply that older patients are more likely than younger patients to get influenza.

Initial demographic data were recorded when patients were assigned to experimental groups. Thereafter, data gathering occurred either when study patients came to the FMC for vaccination or in mid-December, when they were called and interviewed by phone. At that time, patients in the control group were called to determine if they had been vaccinated. They were also reminded of the availability of vaccine. Telephone interviewing was necessary because only 36.6 per cent of patients were vaccinated in the FMC.

Of 395 patients originally selected for study, 88 were unavailable for analysis because they had moved, died, were hospitalized, had withdrawn from the practice or had no phone. Of the remaining 307 patients, data were gathered on 283 (92 per cent response rate). Twenty-four could not be reached despite at least five phone calls, and one refused to be interviewed.

Statistical comparisons were made using Student's t-Test, chi-square Test or Fischer's Exact Test.[22] Linear logistic regression and the method of maximum likelihood[23] were used to adjust for the observed effect of age and prior vaccination status on vaccination rates.

Results

The patients were, as expected, an elderly and primarily female group. Mean age was 66.7 years (standard deviation, 16.4 years), and 67.7 per cent of the patients were female. The overall vaccination rate was 33.6 per cent, compared with a rate of 41.9 per cent for these patients in 1977–1978. Vaccination rates before the 1976 swine flu program had consistently been around 50 per cent.[3]

Only 36.6 per cent of patients were vaccinated at the FMC, and only 20.1 per cent of those vaccinated were vaccinated at a routine appointment. Other vaccination sites included the place of residence (19.6 per cent), the county Public Safety Building (9.7 per cent) and other sites including physicians' offices, fire stations, schools and work places (34.1 per cent). Patients vaccinated outside the FMC generally obtained vaccination free of charge. They were usually told of such options when they called the FMC for their appointment, which perhaps accounts for the high out-of-clinic vaccination rate.

Persons vaccinated had a greater mean age than persons not vaccinated (72.1 ± 10.1 years vs. 63.9 ± 18.2; $t = 4.09$, $p < 0.001$). Patients with pulmonary disease had a significantly lower ($p < 0.05$) vaccination rate (14.3 per cent) than patients with cardiac disease (41.0 per cent) or patients selected for study solely because they were over 65 years old (37.7 per cent). Vaccination rates slightly greater ($0.05 < p < 0.1$) for patients (high utilizers) with more than three clinic visits in the previous six months (41.5 per cent) compared with low utilizers (28.1 per cent), and for patients without a history of adverse reaction to influenza vaccine (50.0 per cent) compared with patients with an adverse reaction to influenza vaccine (31.7 per cent). Vaccination rates did not differ for males and females, or for persons with and without a past history of influenza.

Table 1 shows that the experimental groups did not differ by age, sex ratio, prior history of influenza or adverse reaction to vaccine ($p < 0.05$). Other covariables examined that were distributed similarly among experimental groups were diagnostic classification and clinic utilization rates. However, the Health Belief Model and Personal groups did contain more patients who had vaccinated last year (VLY) or anytime in the past five years than did the Control group or the Neutral group ($0.01 < p < 0.05$). Why randomization failed to distribute these variables more evenly is unexplained and presumably due to chance. Unfortunately, information regarding prior vaccination status was missing for 68 patients (24.0 per cent); a disproportionate share of these patients (34) were in the Control group.

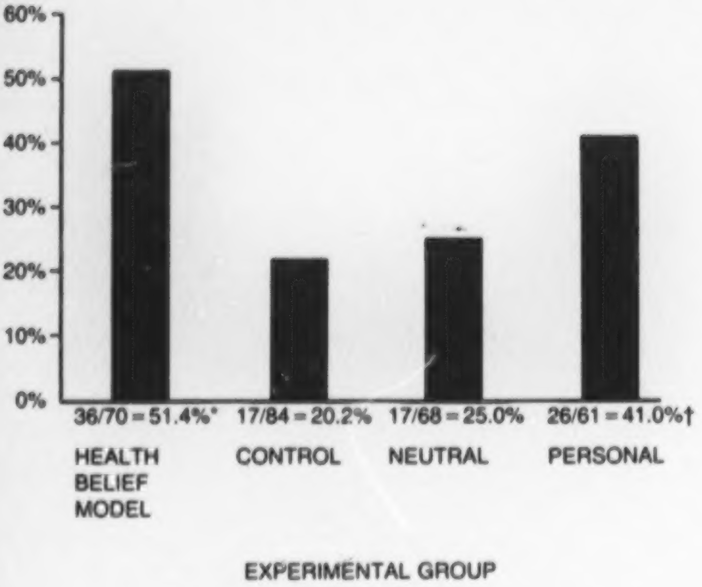
The vaccination rates for each of the four experimental groups are displayed in Fig. 2. Patients receiving the HBM postcard had a higher vaccination rate (51.4 per cent) than control patients (20.2 per cent) and than patients receiving a neutral postcard (25.0 per cent). Patients receiving a personal postcard also had a higher vaccination rate (41.0 per cent) than control patients.

Because the groups differed with respect to past vaccination experience, vaccination rates were computed separately for paients with different vaccination histories (Table 2). The Health Belief Model group consistently had the highest vaccination rate, no matter how the groups were divided. However, smaller and unequal cell sizes resulted in differences that were statistically significant ($0.05 < p < 0.1$) for some comparisons.

Linear logistic regression was then used to test the statistical significance of the variation of vaccination probability with the

TABLE 1. Comparison of Study Groups

	Experimental Group			
	Control	Neutral	Health Belief Model	Personal
Per cent female	65.1	63.2	68.6	75.4
Mean age ± SD (years)	68.0 ± 15.4	63.8 ± 19.6	68.1 ± 14.4	66.5 ± 16.1
History of adverse reactions to vaccine (%)	27.0	19.6	22.6	23.3
History of influenza in past (%)	22.9	25.0	20.7	38.6
Per cent vaccinated last year	34.0	34.5	51.6	51.1
Per cent vaccinated any time in past five years	45.5	44.8	65.6	63.6



*x^2 = 16.40, Control × H.B.M., $p<.001$
x^2 = 10.32, Neutral × H.B.M., $.001<p<.01$
†x^2 = 7.39, Control × Personal, $.01<p<.025$
x^2 = 3.72, Neutral × Personal, $.05<p<.1$

FIG. 2. Vaccination rates for experimental groups.

following variables of interest: age, sex, experimental group, vaccination status last year (VLY), disease classification, history of adverse reactions to influenza vaccine, history of influenza in the past and utilization status. Age, vaccination status last year and experimental group were the only variables that were found to have a significant effect on the likelihood of being vaccinated. The method of maximum likelihood

TABLE 2. Vaccination Rates and Prior Vaccination Experience

	Experimental Group				
	Control	Neutral	Health Belief Model	Personal	Combined
Patients vaccinated anytime in the past five years	45.0% (9/20)*	53.8% (14/26)	77.5% (31/40)*	57.1% (16/28)	61.4% (70/114)
Patients vaccinated last year (1977–78)	36.2% (9/16)†	63.2% (12/19)	81.3% (26/32)†	66.7% (16/24)	69.23% (63/91)
Patients not vaccinated last year (1977–78)	9.1% (3/33)‡	13.2% (5/38)	26.7% (8/30)‡	13.0% (3/23)	15.3% (19/124)

* $x^2 = 6.32$ (Control × Health Belief Model); $0.01 < p < 0.025$.
† $x^2 = 3.38$ (Control × Health Belief Model); $0.05 < p < 0.10$.
‡ $p = .065$ by Fischer's exact test (Control × Health Belief Model).

was used to obtain estimates of the parameters μ, α, and β in the following model:

$$\ln (p/l\text{-}p) = \mu + \alpha A + \beta x + \gamma_i$$

where p = probability of being vaccinated; A = patient age, x = 1 if the patient was vaccinated last year and 0 otherwise; γ_i takes on different values for the different experimental groups. The likelihood ratio criterion was used to determine the importance of the different variables with respect to being vaccinated.

After adjusting for age and VLY, the hypothesis that there is no difference in the vaccination rate among the four groups receiving either no postcard or one of the three types of postcards was tested. The value for $-2 \ln \Lambda$ was 6.5, (Λ is the likelihood ratio statistic and compares the model without the term for postcards to the model including that term). This statistic is approximately χ^2 and on 3 degrees of freedom has a $p < 0.10$ ($n = 213$). Considering only those patients who did receive postcards, the test of the hypothesis that there is no difference in the vaccination rate among groups receiving different postcards resulted in a $-2 \ln \Lambda$ of 4.5 on 2 degrees of freedom ($p < 0.01$, $n = 164$). Table 3 displays the results of the tests of the hypotheses of no difference in the vaccination rates between the two groups listed after adjusting for age and VLY. The HBM appears to be effective compared with no postcard or the neutral postcard.

Based on our data and the log-linear regression model, we obtained the fitted value for p = probability of being vaccinated for a 65-year-old as a function of vaccination history and postcard group. The data presented in Table 4 offer an estimate of the magnitude of differences between experimental groups.

Unadjusted subset analyses included patients (n = 149) who stated that they had never had influenza. The rates for both the Health Belief Model (53.4 per cent) and Personal (48.1 per cent) groups were significantly higher than for the Control group (21.6 per cent, $p < 0.05$). The Neutral group's rate was mid-range (33.3 per cent). The increased vaccination rate for patients receiving the HBM card was observed for new patients who had enrolled in the clinic within the previous year, as well as for high and low clinic utilizers.

TABLE 3. Results of Tests of the Hypotheses That Vaccination Rates Are Not Different After Adjusting For Age and Vaccination Status

Test	n	$-2 \ln \Lambda$	df	p
HBM vs. Personal	108	2.0	1	< 0.25
HBM vs. Control	111	5.1	1	< 0.025
HBM vs. Neutral	118	4.4	1	< 0.05
Personal vs. Neutral	102	0.45	1	
Personal vs. Control	95	0.58	1	
Any postcard vs. Control	213	1.8	1	< 0.25

Discussion

Physicians and dentists commonly use postcards to remind patients of appointments. Studies have consistently demonstrated that postcards do lower broken appointment rates.[19-21]

The elements of HBM have been shown to correlate with a variety of preventive health behaviors, including influenza immunization. Our study tests the hypothesis that the combination of a cue postcard and a message emphasizing the elements of the HBM are more effective than no postcard or than a postcard containing what we have called a neutral massage, simply announcing the availability of vaccine. The HBM postcard was more effective than either no postcard or the neutral postcard and remained so after adjusting for age and prior vaccination history. Furthermore, subgroup analysis showed that this effect persisted without exception in

TABLE 4. Estimates for p = Probability of Being Vaccinated

Age = 65
(All four experimental groups)

	Not Vaccinated Last Year	Vaccinated Last Year
No postcard	0.10	0.57
Neutral	0.11	0.60
HBM	0.24	0.80
Personal	0.14	0.66

Age = 65
(HBM vs. Neutral Postcard)

	Not Vaccinated Last Year	Vaccinated Last Year
Neutral	0.12	0.61
HBM	0.26	0.80

Age = 65
(HBM vs. No Postcard)

	Not Vaccinated Last Year	Vaccinated Last Year
No postcard	0.10	0.59
HBM	0.25	0.81

the various subgroups analyzed. The consistency of the increased vaccination rates in the group receiving the HBM postcard suggests that the HBM postcard is more effective than a neutral or no postcard.

Studies of vaccination and other preventive health behaviors have also shown that physician advice correlates with compliance.[16] Nonetheless, a personal postcard, designed to provide a message from the patient's personal physician, although effective overall, was not more effective after adjusting for age and prior vaccination experience. This observation does not argue that physician advice is not effective, but that the personal postcard substitute for direct physician advice used in this study was not demonstrably more effective than no postcard or the neutral postcard. Given the methodologic problems we encountered, we hope this type of cue will be restudied.

The major methodologic problem in this experimental study is the failure of the randomization to distribute prior vaccination experiences equally in the four groups. This failure reemphasizes the need to collect data on important covariables even in randomized clinical trials. Such a failure was presumably due to chance but raises the possibility that other, undetected covariables (socioeconomic status, prior health beliefs, etc.) might not have been distributed randomly. Thus, we qualify our conclusions and are especially hopeful that our study will be repeated with a more effective randomization of covariables. A nonuniversity practice setting for such a trial would be particularly appropriate, since a trial in this setting would also allow assessment of the generalizability of our results.

Another methodologic concern is the fact that we relied on patient self-report of vaccination status rather than on a more objective method. The use of telephone follow-up was necessary because almost two thirds of our patients were vaccinated outside the FMC, at such a variety of sites that objective validation of vaccination status was impossible. Although not ideal, we used self-report in our earlier study of vaccination compliance[3] and found a greater than 90 per cent concordance between self-reported vaccination status and the FMC records, which suggests that the method is fairly reliable, at least in this setting.

We know of no other study that has tested various postcard reminders in this fashion. The finding that the HBM postcard was more effective is of some theoretical interest to scholars interested in the general development of the model. The general observation that retrospective studies find higher correlations between the HBM than prospective studies[17,24,25] and the results of Cummings et al.[16] showing that the HBM variables influence behavior through behavioral intention suggested that health beliefs change and may be changeable. The postcard emphasizing those elements of the HBM that were correlated with vaccination in our earlier studies might have changed persons' "subjective probabilities"[9] or expectations that a particular action will produce a desired outcome, perhaps by interacting with or reinforcing certain preexistent health beliefs. The change was presumably great enough to promote the desired behavior.

The results support the notion that the HBM is effective as a source of material for health-related advertising campaigns. In addition, for the practitioner, the observed effectiveness of the HBM postcard suggests that the HBM may be a useful approach in counseling designed to change unhealthy behaviors. Inui et al.[26] included the HBM in physician tutorials designed to improve physicians' effectiveness as managers and educators of patients with essential hypertension. Patients of tutored physicians were more compliant with drug regimens and had better control of blood pressures than patients of untutored physicians.

Regarding influenza vaccination, others have commented on the disappointingly poor compliance rate.[27-29] The progressive fall in overall vaccination rate observed in the FMC patients following the swine flu scare suggests that negative attitudes of the public toward influenza vaccines may be more prevalent in the post-swine-flu era. These observations indicate a need for techniques such as effective postcard reminders to promote vaccination if vaccination of a substantial number of high-risk patients is desirable. It must be acknowledged that Sabin[30] has challenged the program of annual influenza immunization because of declining excess mortality associated with recent influenza epidemics and the fact that most bed disability from clinical influenza is not caused by influenza viruses. However, Sabin's analytic methods have been challenged,[31] and many authorities do not agree with Sabin's recommendations.[27,28,32] The Public Health Service Advisory Committee on Immunization Practices continue to recommend yearly vaccination in spite of Sabin's conclusions.

The recent availability of pneumococcal vaccine has also focused increased attention on the general phenomenon of low utilization of vaccines designed to benefit special target populations.[33] Our study suggests that a HBM postcard is one effective way for practitioners to promote vaccines for high-risk patients in their practice. Reminder postcards can be designed to explain that patients are at high risk (i.e., *susceptible* to a potentially severe disease) for the disease for which an effective low-risk vaccine is available. A health-belief-based reminder postcard promoting influenza vaccination among high risk patients would probably be effective in promoting other vaccines like pneumococcal vaccine.

References

1. Influenza vaccine: preliminary statement. Morbidity Mortality Weekly Rep 1978;27:205.

2. Influenza vaccine: recommendation of the Public Health Services Advisory Committee on Immunization Practices. Morbidity Mortality Weekly Rep 1975;24:197.

3. Larson EB, Olsen E, Cole W. The relationship of health beliefs and a postcard reminder to influenza vaccination. J Fam Med 1979;8:1207.

4. Kasl SV, Cobb S. Health behavior, illness behavior, and sick role behavior. Arch Environ Health 1966;12:246.

5. Andersen R. A Behavioral Model of Families' Use of Health Services. Center for Health Administration Studies, University of Chicago, Research Series 25, 1968.

6. Anderson JG, Bartkus DE. Choice of medical care: a behavioral model of health and illness behavior. J Health Soc Behav 1973;14:348.

7. Fabrega H. Toward a model of illness behavior. Med Care 1973;11:470.

8. Becker M, ed. The health belief model and personal health behavior. Health Educ Monogr 1974;2:236.

9. Maiman LA, Becker MH. The health belief model: origins and correlates in psychological theory. Health Educ Monogr 1974;2:336

10. Kegeles SS. Some motives for seeking preventive dental care. J Am Dent Assoc 1963;67:90.

11. Heinzelmann F. Factors in prophylaxis behavior in treating rheumatic fever: an exploratory study. J Health Human Behav 1960;1:183.

12. Hochbaum GM. Public Participation in Medical Screen Programs: A Socio-Psychological Study. Washington, D.C., United States Government Printing Office, 1958. Public Health Service, Public Health Service Publication No. 572.

13. Kirscht JP, Becker MH, Eveland JP. Psychological and social factors as predictors of medical behavior. Med Care 1976;5:422.

14. Becker MH, Drachman RH, Kirscht JP. Predicting mothers' compliance with pediatric medical regimens. J Perdiatr 1972;81:843.

15. Rundall TG, Wheeler JRC. Factors associated with utilization of the swine flu vaccination program among seniors in Tompkins County. Med Care 1979;17:191.

16. Cummings KM, Jette AM, Brock BM, Haefner DP. Psychosocial determinants of immunization behavior in a swine influenza campaign. Med Care 1979;17:639.

17. Sackett DL. Patients and therapies: getting the two together. N Engl J Med 1978;298:278.

18. Taylor DW. A test of the health belief model in hypertension. In: Haynes RD, Sackett DL, Taylor DW, eds. Compliance in health care. Baltimore: Johns Hopkins University Press, 1979;105-109.

19. Schroeder SA. Lowering broken appointment rates at a medical clinic. Med Care 1973;11:75.

20. Shepard DS, Mosely TAE. Mailed versus telephoned appointment reminder to reduce broken appointments in a hospital outpatient department. Med Care 1976;14:268.

21. Nazarian LF, Mechaber J, Charney E. Effect of a mailed appointment reminder on appointment keeping. Pediatrics 1974;53:349.

22. Armitage P. Statistical methods in medical research. New York: Wiley, 1971.

23. Cox DR. Analysis of binary data. London: Wiley, 1970;14-29.

24. Becker MH, Maiman LA, Kirscht JP, et al. Patient perceptions and compliance: recent studies of the health belief model. In: Haynes RB, Sackett DL, Taylor DW, eds. Compliance in health care. Baltimore: Johns Hopkins University Press, 1979;78.

25. Haynes RB. A critical review of the "determinants" of patient compliance with therapeutic regimens. In: Sackett DL, Haynes RB, eds. Compliance with therapeutic regimens. Baltimore: Johns Hopkins University Press, 1978;35.

26. Inui TS, Yourtee EL, Williamson JW. Improved outcomes in hypertension after physician tutorials: a controlled trial. Ann Intern Med 1976;84:646.

27. Seal JR, Sencer DJ, Meyer HM. A status report on national immunization against influenza. J Infect Dis 1976;133:715.

28. Fedson DS. Influenza: the continuing need and justification for immunization. Primary Care 1977;4:761.

29. Koplin AN, Francis BJ, Martin RJ, et al. Administrative costs of the influenza control program of 1976–1977 in Illinois. Med Care 1979;17:201.

30. Sabin AB. Mortality from pneumonia and risk conditions during influenza epidemics: high influenza morbidity during nonepidemic years. JAMA 1977;237:2823.

31. Gregg MB, Bergman DJ, O'Brian RJ, Miller JD. Influenza-related mortality. JAMA 1978;239:115.

32. Salk J, Salk D. Control of influenza and poliomyelitis with killed virus vaccines. Science 1977;195:834.

33. Pantell RH, Steward TJ. The pneumococcal vaccine: immunization at a crossroad. JAMA 1979;241:2272.

BEST AVAILABLE COPY

C5170-1 7/15/82

Medical Care

July 1982
Vol. 20 No. 7
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project

Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems

Effect of a WIC Program on Children's Clinic Activity in a Local Health Department

Closure Among U.S. Community Hospitals, 1976–1980: A Description and a Predictive Model

Self-Care Education: Impact on HMO Costs

Reimbursing for Residency Training: How Many Times?

Procedures and Discrepancies of Blood Pressure Measurements in Two Community Health Centers

Effectiveness of an Ambulatory Care Telephone Service in Reducing Drop-In Visits and Improving Satisfaction With Care

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

**Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington**

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. **Address for subscription information, orders or changes of address:** (except Japan) 2350 Virginia Avenue, Hagerstown, MD 21740. In Japan, contact Woodbell Scope, 11-11, Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. **Annual subscription rates:** U.S. $45.00 individual, $50.00 institution; all other countries except Japan, $55.00 individual, $60.00 institution; special student rate, $27.00. Single copies $6.00. Copies will be replaced without charge if the publisher receives a request within 60 days of the mailing date in the U.S. or within 3 months in all other countries. Printed in U.S.A. © Copyright 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, MD, and at additional mailing offices.

POSTMASTER: Send address changes to Medical Care, 2350 Virginia Avenue, Hagerstown, MD 21740.

July 1982
Volume 20, Number 7

CONTENTS

POLICY FORUM

On "A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project." *David Banta* 661

ORIGINAL ARTICLES

A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project. *Thomas E. Kottke, Pekka Puska, Roger Feldman, Jukka T. Salonen and Jaakko Tuomilehto* 663

Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems. *Paul A. Nutting, Barton R. Burkhalter, Duane L. Dietrick and Edward F. Helmick* 676

Effect of a WIC Program on Children's Clinic Activity in a Local Health Department. *Jonathan B. Kotch and David Whiteman* 691

Closure Among U.S. Community Hospitals, 1976–1980: A Description and a Predictive Model. *Ross M. Mullner, Calvin S. Byre, Paul S. Levy and Joseph D. Kubal* 699

Self-Care Education: Impact on HMO Costs. *Donald W. Kemper* 710

Reimbursing for Residency Training: How Many Times? *Mohan L. Garg, Mounir Elkhatib, Warren M. Kleinberg and Jack L. Mulligan* 719

Procedures and Discrepancies of Blood Pressure Measurements in Two Community Health Centers. *Larry W. Scherwitz, Lynn A. Evans, Deborah J. Hennrikus and Carlos Vallbona* 727

Effectiveness of an Ambulatory Care Telephone Service in Reducing Drop-In Visits and Improving Satisfaction With Care. *Clarence F. Stirewalt, Margaret W. Linn, Gustavo Godoy, Felipe Knopka and Bernard S. Linn* 739

BOOK REVIEWS

Antitrust and the Health Care Provider, by Martin J. Thompson. *Michael R. Pollard* 749

Staffing Primary Care in 1990: Physician Replacement and Cost Savings, edited by Jane Cassels Record. *James E. C. Walker* 750

Politics and the Expanding Physician Supply, by Michael L. Millman. *Eugene Feingold* 751

Managing Change and Collaboration in the Health System: The Paradigm Approach, by Alan Sheldon. *Howard S. Zuckerman* 753

MEDICAL CARE
July 1982, Vol. XX, No. 7

Original Articles

A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project

THOMAS E. KOTTKE, M.D.,* PEKKA PUSKA, M.D.,†‡
ROGER FELDMAN, PH.D.,§ JUKKA T. SALONEN, M.D.,‡
AND JAAKKO TUOMILEHTO, M.D.‡

Permanent and temporary disability pension award data and survey data were used to estimate the lower bound of earning losses related to cardiovascular disease during a community-based cardiovascular disease prevention program in Eastern Finland. Earning losses due to death totaled $39.94 million but were not affected by the project. Earning losses due to permanent disability totalled $29.01 million and were $4.25 million less than expected ($p < 0.025$). Earning losses due to temporary disability were $10.91 million and were not affected by the project. Total earning losses attributable to cardiovascular disease during the project period were $79.86 million, and total decline in earning losses was $4.25 million. Project implementation costs were less than one per cent of total earning losses and were equal to approximately 17 per cent of the decline in these losses attributable to the project. The findings suggest that community-based heart disease prevention programs have the potential of more than paying for themselves through an associated decline in lost earnings.

CARDIOVASCULAR DISEASE (CVD) places a burden on the community by interfering with a person's ability to work. Rice[1] estimated that CVD cost the United States $6.4 billion in 1963, and Cooper and Rice[2] estimated that CVD cost the United States over $40 billion in 1972. They concluded that indirect costs—lost work or livelihood—accounted for 64 per cent of the total costs in 1967 and 73 per cent of the total costs in 1972. Acton[3] estimated that deaths due to heart attacks alone cost the United States $12 billion in lost earnings in 1967. In Finland in 1972, CVD cost $26 million in disability payments (approximately 25 per cent of all disability payments) and an estimated $100 million in lost earnings.[4] In 1973, CVD accounted for almost 15 per cent of all sickness benefit days in England and Wales.[5] In the same year, the proportion of these diseases at-

---

* Laboratory of Physiological Hygiene and § Center for Health Services Research, School of Public Health, University of Minnesota, Minneapolis, Minnesota.

† Epidemiological Research Unit, National Public Health Laboratory of Finland, Helsinki, Finland.

‡ Coordinating Center of the North Karelia Project, Research Institute of Public Health, University of Kuopio, Kuopio, Finland.

Supported in part by NIH training grant 5T32 HL-07036 and by the Coordinating Center of the North Karelia Project.

Address for reprints: Thomas E. Kottke, M.D., Laboratory of Physiological Hygiene, Stadium Gate 27, 611 Beacon Street SE, Minneapolis, MN 55455.

0025-7079/82/0700/0663/$01.15 © J. B. Lippincott Co.

tributable to modifiable risk factors cost Canada more than $500 million; almost 90 per cent of the cost was due to lost income.[6] However, Weinstein and Stason,[7] in their theoretic cost-benefit analysis, suggested that CVD prevention programs based on hypertension treatment cannot recover their own costs, mostly because they projected the beneficial effects to occur far in the future, while treatment costs begin to accrue immediately.

An intervention program may reduce work loss by decreasing disease prevalence or incidence. Conversely, increased case-finding or increased awareness and apprehension about heart disease might increase work loss and the demand for disability pensions and related social benefits. For these reasons, and because cost-effectiveness has become an issue in public health programs, as part of the North Karelia Project evaluation two questions about circulatory diseases and lost livelihood were addressed: 1) How much did heart and circulatory diseases cost North Karelia in lost earnings during the years of the North Karelia Project? and 2) Were there cost changes associated with the Project?

Project Background

The North Karelia Project was started in Finland in 1972 to carry out a comprehensive community program for CVD control. The target group was a county of 180,000 inhabitants with exceptionally high heart disease rates.[8] A quasi-experimental evaluation design with a matched reference area was used to assess the feasibility and effects of this community-based intervention for the five-year period between

TABLE 1. Selected Social and Economic Indicators in North Karelia, the Reference Area and All Finland During the Project Period

Indicator	Year	North Karelia	Reference Area	All Finland
Population (1,000s)	1972	181	253	4639
	1976	177	251	4726
Per cent men	1972	50.0	49.0	48.3
	1976	49.7	48.9	48.3
Proportion of population of working age (aged 15–64)	1972	66.1	65.8	66.9
	1976	67.5	67.0	67.4
Net migration of population aged 15–44 (%)	1972	−2.9	−1.8	+0.1
	1976	−1.1	−0.6	−0.4
Persons per household	1970	3.4	3.4	3.0
	1975	3.0	3.0	2.7
Persons per room	1970	1.3	1.2	1.0
	1975	1.0	1.0	0.9
Per cent of labor force engaged in farming or forestry	1970	39.4	34.3	20.3
	1976	29.0	25.4	15.0
Taxable annual income (Finnmarks/person)	1972	3824	4156	5771
	1976	8999	9640	12631
Per cent of working age population receiving disability pension	1972	11.4	9.6	6.6
	1976	13.0	11.9	8.1
Per cent of population receiving social assistance	1972	6.3	6.2	5.5
	1976	4.9	4.4	4.1

spring 1972 and spring 1977. The reference area was selected because it was the county in Finland that most resembled North Karelia in age structure, risk factor levels, disease incidence rates and socioeconomic characteristics (Table 1). Before the beginning of the intervention and after the end of the pilot period, separate 6.5 per cent samples of the middle-aged populations (born between the years of 1913 and 1947) of North Karelia and the reference area were asked to answer questions regarding beliefs about CVD, risk-factor-related behaviors and socioeconomic status. Participants also had physical examinations. About 91 per cent of the selected sample participated in the surveys (Table 2).

The project intervention goal was to influence disease rates by reducing the population levels of the major cornary heart disease (CHD) risk factors, i.e., smoking, serum cholesterol and blood pressure. Community-based antismoking health education was provided, general dietary changes were promoted to influence serum cholesterol levels, and systematic hypertension detection, treatment and follow-up were organized. The major intervention elements included public health education, organizing individual and group services, promoting environmental changes, training local personnel, and monitoring the changes and program impact.[9] For efficiency and effectiveness, the activities were integrated into the existing community service structure and social organization. The project team helped to plan the activities, supervise their implementation, train other personnel and

TABLE 2. Response Rates to North Karelia Project Population Surveys

	North Karelia		Reference Area	
	Men	Women	Men	Women
1972 Survey				
Original sample (n)	2576	2539	3709	3639
Per cent excluded*	6.6	4.6	5.7	3.5
Purified sample (n)	2407	2423	3499	3510
Nonparticipation (n)				
Temporarily away	65	33	84	69
Unable to answer†	17	14	26	18
Refused	8	3	2	4
Unknown reason	89	66	228	173
Total	179	116	340	264
Participants (n)	2228	2307	3159	3246
Per cent of purified sample	92.6	95.2	90.3	92.5
1977 Survey				
Original sample (n)	2355	2373	3310	3469
Per cent excluded*	1.6	1.3	1.4	0.9
Purified sample (n)	2317	2342	3263	3439
Nonparticipation (n)				
Temporarily away	35	23	45	51
Unable to answer†	30	14	39	33
Refused	23	19	15	29
Unknown reason	227	165	246	168
Total	315	221	345	281
Participants (n)	2002	2121	2918	3158
Per cent of purified sample	85.4	90.6	89.4	91.8

* For reasons of death or emigration.

† For reasons of illness or mental incapacity.

prepare educational materials. The team also catalyzed program development in the community.

A thorough evaluation has demonstrated the success of this program.[9] The program was associated with risk-factor-level declines of 17 per cent among men and 12 per cent among women after adjustment for simultaneous changes in the reference community.[10]

Disease Prevalence and Income and Labor Force Participation Estimates

In order to estimate the effect of CVD on earning losses, the number of workers with CVD, the average income per worker, unemployment rates and labor force participation rates were first estimated. The methods for estimating these rates are presented in this section.

Disease Prevalence Estimation. In this study, heart disease was considered to include stroke, hypertension, congestive heart failure, myocardial infarction, arrhythmia and angina pectoris. The net change in the prevalence of subjects reporting heart disease diagnosed or treated during the previous year was computed by subtracting the change in reporting in the reference area from the change in reporting in North Karelia:

$$\text{net change in CVD prevalence} = (NK_2 - NK_1) - (R_2 - R_1),$$

where NK refers to North Karelia; R refers to the reference area; 1 refers to the baseline survey; and 2 refers to the terminal survey.

The change in disease prevalence was tested for statistical significance with the normal distribution:

$$Z = \frac{\text{net change in CVD prevalence}}{\sqrt{p_{NK}(1 - p_{NK})(1/n_{NK_1} + 1/n_{NK_2}) + p_R(1 - p_R)(1/n_{R_1} - 1/n_{R_2})}}$$

Income Estimation. Mean expected income for individuals free of disease in 1972 and 1977 was calculated from the survey responses to the question "How big was the total income of your household last year, not excluding taxes?" by subjects who defined themselves as main family financial supporters and who denied having any disease treated or diagnosed by a physician within the previous year. Subjects could make one of nine precoded responses about their income; each response, except those for the highest and lowest income, was recoded to be equivalent to the median income in the response range. The highest income response ("over 32,000 Fmk/year") was recoded as "34,000 Fmk/year." The lowest income response ("below 4,000 Fmk/year") was recoded as "2,000 Fmk/year." Annual income estimates were divided by 250 for estimate of daily income.

Employment and Labor Force Participation Rate Estimation. Unemployment would prevent some of those out of work because of disability from woking even if they were not disabled. Mushkin[11] set a now accepted precedent of assuming full employment (e.g., 96 or 97 per cent) as the proportion of disabled who would be working if not disabled. However, Klarman[12] has pointed out that this assumption may overestimate the impact of disease on the whole economy and that the economic value of health reform depends on the rate of economic development. Therefore, because unemployment rates averaged approximately 7 per cent in North Karelia during the project period, income-loss estimations related to permanent disability were deflated by this figure to allow for unemployment. For short-term work losses, it was assumed that the system could not replace the temporarily disabled worker with a healthy one on short notice, so the income loss figures were not deflated.

Even with full employment, not all physically able individuals participate in the labor force. Labor force participation rates for Finland in 1975 (ages 40–44, 0.91; 45–49, 0.92; 50–54, 0.84; 55–59, 0.77; 60–64, 0.52)[13] were used to project labor force participation of the disabled had they not been disabled. Only 9 per cent of the Finnish people over the age of 64 were active in the labor force, possibly because the old age pension replaces close to 100 per cent of the person's previous earnings; it was assumed, therefore, that those over the age of 65 did not contribute to the labor force.

To be conservative in estimating hypothetical labor force participation among disabled in the absence of diability, the lowest labor force participation rate of those without CVD pensions, that of 60–64 year olds, was used, and the 0.52 labor force participation rate was applied to all CVD pensioners regardless of age.

Income Loss Estimates

The data sources for the work loss estimates are presented in Table 3.

Income Loss Due to Death. Income loss due to premature death was defined as:

$$\sum_{t=1}^{65-i} (pS_{it})(pLFP_{it})(E_{it}/1.1^{t})$$

where pS_{it} is the probability that a worker of current age "i" survives for "t" more years; $pLFP_{it}$ is the probability that a worker of current age "i" will be in the labor force "t" years in the future; and $E_{it}/1.1^{t}$ is the annual earnings discounted annually at 10 per cent. In this analysis, "i" was set at 40, 50 and 60.

The age-stratified cause-specific number of deaths in North Karelia during the project period was available through 1975 from national statistics.[14] The numbers of deaths for 1976 and 1977 were estimated from data for 1972 through 1975. The mean age in the 35–44 age group was assumed to be 40, the mean age in the 45–54 age group was assumed to be 50, and the mean age in the 55–64 age group was assumed to be 60. The groups were assumed to have 25, 15 and five remaining working years of life, respectively, if death did not intervene because labor force participation rates after the age of 65 are less than 10 per cent. The probability of survival until the age of 65 was estimated from Finnish life tables for 1975.

Change in death rate was estimated by subtracting the change in the reference area rates from the rate changes in North Karelia:

$$\text{net change in CVD death rates} = (NK_2 - NK_1) - (R_2 - R_1).$$

The statistical significance of the death rate change was tested using the normal

TABLE 3. Data Sources for Estimating Work Losses

Source of Work Loss	Source of Rate Estimation	Specificity of Rate Estimation
Death	Death certificate data	Age- and sex-specific rates
Permanent disability	Pension institute data	Sex-specific and age-adjusted rates
Temporary disability		
Periods lasting ≥ 8 days	Pension institute data	Crude rates
Periods lasting < 8 days	Population survey data	Crude rates

distribution:

$$Z = \frac{\text{net change in CVD death rate}}{\sqrt{p_{NK}(1 - p_{NK})(1/n_{NK_1} + 1/n_{NK_2}) + p_R(1 - p_R)(1/n_{R_1} + 1/n_{R_2})}}.$$

Income Loss Due to Permanent Disability. Although disability pensions are transfer payments, they are a good surrogate measure of disability-related income loss in the Finnish system because they are indexed to predisability income. All residents of Finland aged 15 to 64 years who are premanently unable to work because of disease may apply for a permanent pension.[15] These pensions replace approximately 45 per cent of the person's income. The national Social Insurance Institute administers all disability pensions in Finland, and a pension is awarded only after expert committee review of the patient's request and the treating physician's certification of disability. North Karelia and the reference county are in the same insurance district. Age, sex, disease and county-specific permanent pension data were available from the Social Insurance Institute of Finland for the years from 1968 to 1978 (unpublished data).

The cost of permanent disability pensions was calculated by multiplying the number of CVD-related disability pensions each year by the national average award size for that year. The annual figures were summed for the years 1972 through 1977. The earning loss due to permanent CVD-related disability was estimated by dividing the pension costs by 0.45, the reported average replacement ratio. This figure was multiplied by 0.52, the labor force participation rate in 1975 for the population aged 60–64 years, and deflated by 7 per cent to allow for unemployment effects.

Piecewise regression[18] was used to test for and estimate the size of project effects on permanent pensions. Variables for county, year (linear and quadratic components), and program (the "program" variable was constructed by assigning "0" for all years for the reference area, "1" for North Karelia in 1972, "2" for North Karelia in 1973 and so forth, through "7" for North Karelia in 1978) were regressed on total, CVD-related and non-CVD-related disability pension rates. The data were age adjusted to the 1972 population of Finland, and the analyses were performed separately for men and women. Because all CVD events occurred in persons over 35 years of age and age-specific data were available, accuracy was increased by limiting permanent pension analysis to the 35–64-year-old age group.

Income Loss Due to Temporary Disability

All working-age residents of Finland who are unable to work for periods of eight to 300 consecutive workdays are eligible for "daily allowance" payments.[17] A work year is considered to be made up of 250 workdays. Daily allowance payments pay workers 0.15 per cent of their previous year's annual income for each day that they are unable to work, providing their income falls within a defined range. If their income is below the range, it is replaced at a flat rate. If total income exceeds the range, that part of the income above the ceiling is ignored, and workers receive payment at the rate for the ceiling income. Payments to household heads are supplemented 15 per cent for a spouse and 10 per cent for each child, providing that the total supplements do not exceed 50 per cent of the basic daily allowance rate. If workers are included in the obligatory workers' insurance plan through their employers or if they receive benefits from a private insurance plan, the government reimburses the other plan for its expenses with the same formula as that used for payments to individuals. Replacement ratios are reported to average about 45 per cent of earning losses. As a rule, workers' insurance and private insurance plans do provide coverage at reimbursement levels greater than those of the Social Security Institute. However, these payments are not included in the analysis because they only represent further transfer payments and are not necessary to consider when estimating production losses.

Disease and county-specific data were available from the Social Security Institute of Finland for the years 1972 through 1977 (unpublished data). The cost to the government for daily allowance payments was calculated by summing the CVD-related daily allowance payments each year for the years 1972 through 1977. Income loss due to CVD-related disability periods greater than eight days was estimated by dividing this figure by 0.45, adding the income from eight working days for each daily allowance period initiated and then deflating by 7 per cent to allow for unemployment effects.

A difference in slopes of the pension rates in the two counties (i.e., interaction of year and county) over the five project years was defined as a sign of program effects on daily allowance costs. Dependent variables for the CVD and non-CVD daily allowance costs and daily allowance periods initiated were regressed on independent variables for county, year and county-year interaction. A p value of ≤ 0.05 was required to reject the null hypothesis of no county–year interaction.

Work losses attributable to CVD for periods totaling less than eight consecutive workdays could not be estimated directly. Therefore, they were estimated in the following manner: all survey respondents who were not pensioned, reported missing less than eight days from work or major activity in the previous year, and reported having either no disease or only CVD treated or identified by the physician in the previous year were identified. The estimations were based on this population. The work-loss rate (days/year) in subjects with no diseases was subtracted from the work-loss rate for individuals who reported having only CVD treated or diagnosed by a physician during that year. This can be described arithmetically as:

$$\text{work loss due to CVD} = (NK_{CVDO} - NK_{ND}),$$

where NK_{CVDO} is the days of work lost reported by those who missed less than eight days of work and had only CVD, and NK_{ND} is the days of work lost reported by those who missed less than eight days of work and reported no disease.

The work loss per worker was multiplied by the number of workers in the county with CVD to give the number of work-loss days in the county due to periods lasting less than eight days. The resulting number was multiplied by the calculated daily earnings to calculate the total earning losses in the country.

The calculation for the change in work loss in North Karelia over the project period then appears as:

$$(NK_{CVDO_2} - NK_{ND_2}) - (NK_{CVDO_1} - NK_{ND_1}),$$

where "2" refers to 1977 and "1" refers to 1972.

To adjust for spontaneous changes over time, the change in work loss in the reference area was subtracted from the change in work loss in North Karelia:

$$[(NK_{CVDO_2} - NK_{ND_2}) - (NK_{CVDO_1} - NK_{ND_1})] - [(R_{CVDO_2} - R_{ND_2}) - (R_{CVDO_1} - R_{4ND_1})],$$

where "R" refers to the reference area.

$$Z = \frac{\text{net work loss change}}{\sqrt{S_p^2(1/n_{NK_1} + 1/n_{NK_2} + 1/n_{R_2} + 1/n_{R_1})}},$$

where

$$S_p^2 = \frac{(n_{NK_1} - 1)S^2_{NK_1} + (n_{NK_2} - 1)S^2_{NK_2} + (n_{R_1} - 1)S^2_{R_1} + (n_{R_2} - 1)S^2_{R_2}}{(n_{NK_1} + n_{NK_2} + n_{R_1} + n_{R_2}) - 4}.$$

The normal distribution was used to test for a significant change in the work-loss rate attributable to heart disease in subjects claiming less than eight days lost per year.

Results

Prevalence of CVD in the Labor Force

In 1972 and 1977, the proportion of persons on the North Karelia project surveys who claimed a CVD disease treated or diagnosed by a physician in the previous year averaged 20.5 per cent and increased from 17.8 per cent in 1972 to 23.2 per cent in 1977 in North Karelia. The proportion averaged 19.1 per cent and increased from 15.9 per cent in 1972 to 22.3 per cent in 1977 in the reference area. The net decline in North Karelia of 1.0 per cent was nonsignificant ($Z = 0.78$, $p = 0.29$).

Income Estimations

The mean income among major household providers in North Karelia was $4,031 per year in 1972 and $6,887 per year in 1977, a mean of $5,459 during the project period (all figures have been converted to U.S. dollars at a rate of 4 Fmk = U.S. $1). Assuming a work year of 250 days, these figures are equivalent to $11.04, $18.87 and $14.96 per day, respectively.

Income Loss Due to Death

From the available age-specific data, it was estimated that there were 154 CVD-related deaths among men and 39 CVD-related deaths among women in the 35–44 age group. There were an estimated 478 CVD-related deaths among men and 119 CVD-related deaths among women in the 45–54 age group and 1,084 CVD-related deaths among men and 271 CVD-related deaths among women in the 55–64 age group during the entire five-year project period. The total earnings lost to CVD-related death during the project period were $39.94 million. Although myocardial infarction death rates declined 16 per cent in men and 5 per cent in women, and stroke death rates declined 38 per cent in men and 50 per cent in women, the net changes (adjusted for changes in the reference area) were not significant.[18] Therefore, death-related income loss did not change.

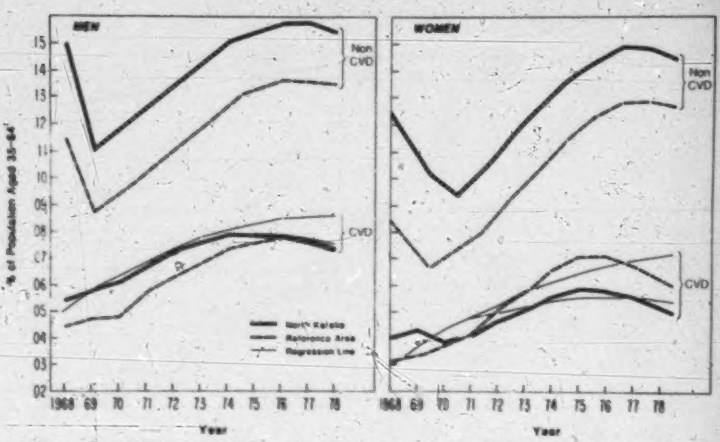
Income Loss Due to Permanent Disability

In North Karelia from 1972 to 1977, government costs for CVD-related permanent disability pensions were $27.00 million. This represents income losses of $29.01 million after adjustment for income replacement ratio (45 per cent), unemployment rates (7 per cent), and labor force participation rates (52 per cent).

Among North Karelian men in 1968, about 5.5 per cent of the population aged 35–64 received a disability pension for CVD. This proportion rose to a peak of 7.9 per cent in 1974. The reference area rates started at 4.5 per cent in 1968 and continued to rise until 1976 so that the rates were similar in the two areas from 1976 to 1978 (Fig. 1). There was a significant project effect on the CVD disability pension rates ($p < 0.025$). The project effect resulted in the observed CVD pension rates averaging 6 per cent less than expected during the overall project period and 10 per cent less than expected in 1977, the last project year. Total observed pension rates tended to be less than expected, but this trend was not significant.

Among North Karelian women, 4 per cent of the population aged 35–64 was pensioned for CVD in 1968. This rate rose to a peak of 5.7 per cent in 1975. The rates in the reference community surpassed the North Karelia rates in 1971 and climbed at a steeper rate until the middle of the decade (Fig. 1). As with men, the project effect on CVD pension rates was significant ($p < 0.025$); this resulted in observed project area rates averaging 13 per cent less than expected during the entire project period and 21 per cent less than expected during the last year of the project. As with North Karelian men, observed total disability pension rates tended to be less than ex-

FIG. 1. Estimated by piecewise regression, the project had significant effects on both men's and women's CVD disability pension rates (both $p < 0.025$). These effects, represented by the areas between the regression lines, yielded a total savings to the government of $3.95 million during the project period. There were no project effects on non-CVD pension rates.



* Age-adjusted to 1972 Finnish population aged 35-64 years.

pected during the project period, but the trend was nonsignificant. Non-CVD pension rates were not affected by the project among either sex.

The decline in pension awards as a result of the deficit in expected CVD pensions was $1.06 million for men and $2.89 million for women—a total of $3.95 million. With adjustment for income replacement ratios, unemployment and labor force participation, this represents a $4.25 million decline in earnings loss.

Income Loss Due to Temporary Disability

Periods Greater than Eight Days Long. The total government costs for daily allowances paid in North Karelia for CVD (ICD VII) was $4.58 million from 1972 through 1977. Using a daily allowance to lost income replacement ratio of 0.45:1 and allowing for 7 per cent unemployment, this represents $9.47 million in lost earnings. Non-CVD daily allowance periods per eligible inhabitant (Fig. 2), and the CVD-related daily allowance periods per eligible inhabitant declined significantly in both North Karelia and the reference area during the project period. Although a trend was present, the rate of decline for CVD-related periods was not significantly greater for North Karelia than for the reference area ($p = 0.11$).

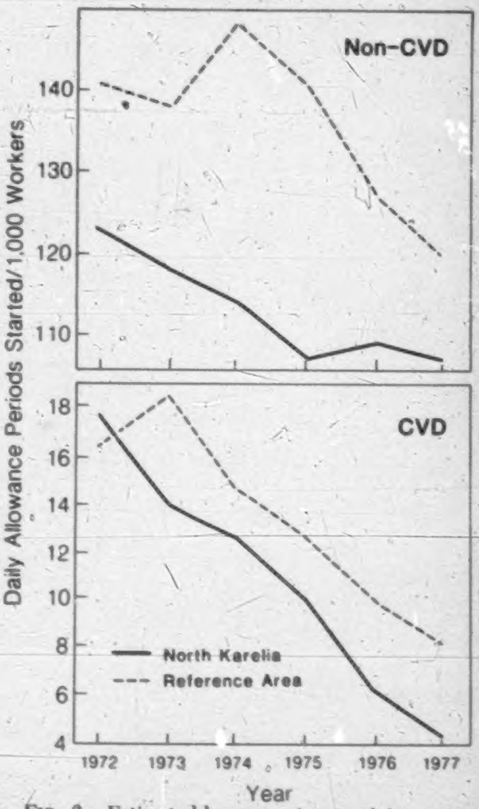


FIG. 2. Estimated by regression, work-loss costs related to temporary disability from all causes and CVD declined significantly during the project period (both $p < 0.05$). Although the decline in work-loss costs from CVD-related temporary disability was greater in North Karelia than in the reference area, this difference was not significant ($p > 0.10$).

TABLE 4. CVD-Related Temporary Disability Work-Loss Costs for North Karelia: 1972 to 1977 (million U.S.$)

	Loss*
Periods lasting >8 days	
Government reimbursed (8 to 300 workdays)	9.47
Deductible period (first 7 workdays)	0.85
Periods lasting <8 days	0.59
Total	10.91

* There were no significant changes in short-term work-loss costs.

There were 7,611 CVD-related daily allowance periods started in North Karelia from 1972 to 1977. These periods were associated with 60,888 work-loss days that were required for eligibility for daily allowance payments but not reimbursed by the system. After adjustment of unemployment, income losses for these periods were estimated to be $0.85 million.

Periods Less than Eight Days Long. In North Karelians who reported CVD and missed less than eight days from work in the year prior to the 1972 survey, work loss attributable to CVD was 1.3 days per 1,000 workdays. The 1977 figure was 0.8 day lost per 1,000 workdays. In reference area workers with CVD who missed fewer than eight days in the previous year, between 1972 and 1977 there was an increase from a negative 0.1 per 1,000 workdays to 0.4 per 1,000 workdays lost that were attributable to CVD. The resulting net decline in CVD-attributable work loss for periods less than eight working days in North Karelia of 1.0 day per 1,000 was nonsignificant ($Z = 0.035$, $p > 0.25$).

Subjects claiming CVD in North Karelia during the project period averaged 20.5 per cent of the nonpensioned labor force (25,000 workers). This rate, combined with a mean work-loss rate attributable to CVD of 1.05 days per 1,000 workdays for periods of less than eight days per year in a work year of 250 workdays, gave a resulting work loss in North Karelia of 6,563 days per year, accrued in periods less than eight workdays long. This category of income loss totaled $589,095 for 1972 through 1977. Adding $589,095 in lost earnings for the illness periods of less than eight days to the income losses for periods of greater than eight days yielded an estimate of $10.91 million in income lost to temporary disability from CVD during the years 1972 through 1977 (Table 4).

TABLE 5. CVD-Related Disability Work-Loss Costs for North Karelia: 1972 to 1977 (million U.S.$)

Loss Area	Loss	Change in Loss
Death	39.94*	Not significant
Permanent disability	29.01	−4.25 ($p < 0.025$)
Temporary disability	10.91	Not significant
Total	79.86	−4.25

* Discounted at 10 per cent to present value dollars.

Total Income Loss

Death-related income losses of $39.94 million were combined with permanent disability-related income losses of $29.01 million and $10.91 million for temporary disability-related income losses to arrive at a figure of $79.86 million in income loss due to CVD during the project period (Table 5). This amount is $4.25 million less than would have been expected without the intervention project.

Discussion

Dollar assessments are one important measure of health benefits because, at some point in policy development, these benefits will be related to the public and private costs of bringing them about. The criterion of economic efficiency is given high priority by some policy-makers and their most influential constituencies.[19]

For this reason, the intention of this article has been to develop the truest possible estimation of CVD-related labor market losses during the North Karelia Project.

Therefore, this article does not present an estimation of total costs, and the reader should be aware that the very important contributions of those who do not directly participate in the labor force, for example, housewives, have not been counted. There has also been no attempt to account for intangible losses like pain, suffering and anxiety by assigning them a shadow price and including these estimations in the calculations.

Although it was more than four million dollars, the difference between the observed and expected income losses in North Karelia could possibly be attributed to processes other than the morbidity changes resulting from the North Karelia Project. For example, administrative decisions might have caused the observed differences. For many years, because of the social aspects of pensions, award criteria might have been lenient in economically distressed areas such as North Karelia. In the early to the mid-1970s, however, the government made an effort to standardize award criteria. Even if this effort had not been present, both North Karelia and the reference area are in the same sickness insurance district, so applications would have received similar treatment in both counties.

Another possible explanation is that migration caused the observed changes. The difference of 1,200 pensioners between the observed and expected numbers in North Karelia was only 2 per cent of the total CVD pensioners in Finland. However, it is not believed that migration explains the differences for two reasons: first, the standardized mortality rate for CVD death was only 1.48 : 1.00, and the ratio of pension awards for North Karelia compared with those for Finland in 1972 was 2 : 1. Criteria for pension awards were at least as lenient in North Karelia as anywhere else in Finland, so the pension seeker had no reason to move from North Karelia to obtain a pension. Second, the emigration from North Karelia that did occur was by young, healthy people looking for jobs in Helsinki and other cities in the south.[20] Older Finns, with their strong family attachments, would rarely leave their home area.

Economic trends could also have confounded the pension rates. In 1973, with the recession following the Arab oil embargo, Finnish exports had difficulties on foreign markets, and North Karelia suffered the most of any Finnish region. However, economic recession drives pension demand up, and the economic changes during the period would have increased, not decreased, pension demand in North Karelia.

The survey data gathered in North Karelia and the reference area also may not exactly reflect the true situation, and the methods for estimating CVD-attributable income loss for periods of less than eight days are far from perfect. Subjects have a difficult time remembering their past behavior, and responses on questionnaires about diseases and diagnoses are also subject to error. Estimates of individual income may be inaccurate because North Karelia is a rural area, and in many cases income cannot be disaggregated below the level of the family unit. Even so, because of the minor contribution of short work-loss periods to the total work-loss burden of CVD, inaccurate estimation could not have significantly influenced the results.

Conversely, to the extent that the findings of this study parallel the changes observed in morbidity, they may be interpreted as resulting from them. First, the rates for stroke declined 38 per cent for men and 50 per cent for women in North Karelia between 1972 and 1977.[18] The myocardial infarction rates declined by 16 per cent in men and 5 per cent in women in the same period. Treatment of hypertension increased equally in both counties, but the hypertensive control rate was significantly greater in North Karelia than in the reference area.[21] There was a concomitant leveling off of the population requiring drugs for congestive heart failure associated with this better control of blood pressure in North Karelia, but not in the reference area.[9] This supports the contention that the North Karelia population was relatively healthier at the end of the period than at the beginning and that the declining pension rates were due to improved health.

This study has avoided the difficulties of extrapolating future trends in disease by limiting its analysis to the five-year pilot project period. Because the changes in health behavior continue spontaneously in North Karelia at this time, however, it is believed that the declines in disability will persist at a steady rate or may even decrease further. If the differences in permanent disability were to persist in a steady state at the rate found at the end of the project, the income savings during any given period would be half again as great as those presented here. The ten-year trends will be analyzed and reported after the population is resurveyed in spring 1982.

Beyond earning losses, disability also represents opportunity costs for a community. Although the healthy worker produces while consuming and the worker who has died neither consumes nor produces, the disabled worker consumes without producing. The "Disability '72" study in the United States found that "chronic poor health impoverishes not only those it afflicts, but also those living with them"[22] because resources normally used for other purposes must be diverted to support the disabled.

Disability also causes social disorganization, and the disabled suffer from a poor self-image. Franklin[22] found that "more marriages end through divorce or separation for the disabled than for the nondisabled," and that the families of the disabled usually contract the scope of their activities to manage with the disabled person.

There are those who hesitate to examine earning losses for fear of being forced into recommending programs that focus on those with the highest per capita earnings. However, if earnings are viewed within the context of the social system, two points become apparent: first, laborers in most technological societies have gained the right of passing health care and disability costs on to their employers. It is this right that has prompted management in the U.S.[23] to seek methods of containing health care costs, and, in at least one state, management, charitable foundations and organized medicine have joined their efforts.[24] The worker who requires less health care costs management less money than the worker who requires more.

Second, although the production of any given worker may not be distinguishable from the production of another worker, the wage earner is a necessary component of technological societies. The effectiveness of the labor strike is adequate testimony to this fact. Because an efficiently working social system requires that all components be in good condition, there is little to gain by maximizing the health of the managers while neglecting the health of the wage earners.

Finally, as was done in Finland, policymakers can simply decide that preventive medicine programs must include all people, and they can adopt programs that attempt to influence the health of the whole population. Besides being equitable, it is believed that this approach may also be the most efficient public health strategy. To shun the analysis of changes in earning losses associated with public health programs during periods of social program retrenchment is to ignore some of the most powerful data supporting the efficacy and cost-effectiveness of the public health effort.

A comparison of the relative magnitudes of the project costs, the health services costs and the work-loss costs is revealing. It was previously reported that project costs were $1.02 milion for evaluation and $730,000 for implementation.[9] Although the trend analysis has not yet been completed, direct community costs (health services) totaled $13.86 million for drugs,

$10.10 million for hospital care of stroke and myocardial infarction, and $20.09 million for primary care—$44.05 million in all.

The calculations in this study represent the lower bound of CVD-related work-loss costs and only a proportion of total CVD-related costs. The work-loss costs were about 100 times the project implementation costs, and the decline in work-loss costs alone was more than six times the project implementation costs. The savings to the government in reduced pension costs alone were five times the project implementation costs. The middle or upper bound estimation would make savings in income losses even larger relative to program costs. These figures strongly suggest that the potential returns in earning capability alone merit investment in heart disease prevention programs.

References

1. Rice DP. Estimating the cost of illness. Am J Public Health 1967;57:424.

2. Cooper BS, Rice DP. The economic cost of illness revisited. Soc Secur Bull 1976;39:21.

3. Acton JP. Evaluating Public Programs to Save Lives: The Case of Heart Attacks. Santa Monica: The Rand Corporation, 1973.

4. Social Insurance Institute of Finland. Statistical Yearbook of the Social Insurance Institution of Finland. Helsinki: Social Insurance Institute of Finland, 1972. (Series T1: number 8.)

5. Black DAK, Pole JD. Priorities in biomedical research: indices of Burden. Br J Prev Soc Med 1975;29:222.

6. Smith ESO. The economic impact of preventable ischemic heart disease. Can Med Assoc J 1977;117:507.

7. Weinstein M, Stason W. Hypertension: a policy perspective. Cambridge, Mass.: Harvard, 1976.

8. Puska P, Mustaniemi H. Incidence and presentation of acute myocardial infarction in North Karelia, Finland. Acta Med Scan 1975;197:211.

9. Puska P, Tuomilehto J, Salonen J, et al. The North Karelia Project: Evaluation of a Comprehensive Community Programme for Control of Cardiovascular Diseases from 1972–1977 in North Karelia, Finland. Copenhagen: WHO Regional Office for Europe, 1981.

10. Puska P, Tuomilehto J, Salonen J, et al. Changes in coronary risk factors during comprehensive five-year community programme to control cardiovascular diseases (North Karelia Project). Br Med J 1979;2:1173.

11. Mushkin SJ, Collings Fd'A. Economic costs of disease and injury. Public Health Rep 1959;74:795.

12. Klarman HE. Economics of Health. New York: Columbia University Press, 1965;168.

13. Central Statistical Office of Finland. Statistical Yearbook of Finland—1979. Helsinki: Central Statistical Office of Finland, 1980;50-51.

14. Central Statistical Office of Finland. Causes of Death in Finland, VI B:128 to VI B:131. Helsinki: Central Statistical Office of Finland, 1975–1979.

15. Social Insurance Institute of Finland. Statistical Yearbook of the Social Insurance Institution of Finland. Helsinki: Social Insurance Institution of Finland, 1971;41, 194. (Series T1: number 7.)

16. Neter J, Wasserman W. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Homewood, Ill.: Richard D. Irwin, Inc, 1974;313.

17. Social Insurance Institute of Finland. Statistical Yearbook of the Social Insurance Institution of Finland. Helsinki: Social Insurance Institute of Finland, 1971;236. (Series T1: number 7.)

18. Salonen JT, Puska P, Mustaniemi H. Changes in morbidity and mortality during a comprehenisve community programme to control cardiovascular diseases during 1972-77 in North Karelia. Br Med J 1979;2:1178.

19. Milio N. Promoting Health Through Public Policy. Philadelphia: FA Davis, 1981;216.

20. Landers PS. In the Shadow of the Factory: Social Change in a Finnish Community. New York: John Wiley and Sons, 1976.

21. Tuomilehto J, Nissinen A, Salonen JT, Kottke TE, Puska P. Community programme for control of hypertension in North Karelia, Finland. Lancet 1980;ii:900.

22. Franklin PA. Impact of Disability on the Family Structure. Social Security Bulletin, May 1977;3.

23. Containing the cost of employee health plans. Business Week, May 30, 1977;245:74.

24. Minnesota Coalition on Health Care Costs, MC . . . health care providers, consumers, payers and educators working with business, government and labor toward cost-effective, quality health services. Minneapolis: Minnesota Coalition on Health Care Costs, undated.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems

PAUL A. NUTTING, M.D.,* BARTON R. BURKHALTER, PH.D.,†
DUANE L. DIETRICK, PH.D.,‡ AND EDWARD F. HELMICK, MPH§

The performance of 11 medical care systems of varying size and payment mechanisms (consisting of six government owned and operated Indian Health Service units, three fee-for-service private practices and two HMOs) was studied. Performance was defined as the percentage of consumers in need of care who received adequate care according to predetermined standards for the process of care for various functions (i.e., prevention, screening, treatment, follow-up) and various health conditions (i.e., prenatal and infant care, hypertension, anemia, UTI). Size was found to have a strong negative relationship to the quality of treatment and follow-up care, but payment mechanism did not show such a relationship. Neither size nor payment mechanism was significantly related to prevention performance. In screening, the results depended on the health condition: size was inversely related to performance of screening for hypertension; HMOs performed significantly better in screening for prenatal anemia; and neither size nor payment mechanism was related to performance of screening for infant anemia.

DESPITE EXTENSIVE STUDY, there is not yet consensus on how the many aspects of organizational structure influence organizational performance. In the health field, much recent attention has focused on the relation of payment mechanisms to system performance, but in spite of convincing arguments that financial incentives should affect performance in health care delivery systems,[1] studies comparing fee-for-service and prepaid payment programs have produced inconclusive results on most issues,[2–9] at least in part due to methodologic problems.[9] More attention has been given to utilization and cost containment than to quality of care, and to inpatient care than to outpatient care.[8] The increased preventive care widely observed in health maintenance organizations (HMOs) is due to the prepayment mechanism that provides a low or zero marginal cost to the user, rather than to a health maintenance philosophy.[10]

Economists have investigated the relationship between firm size (an aspect of structure) and profitability in economy-of-scale studies. Some have found that larger size is generally correlated with greater profitability,[11,12] but many investigators have not found significant correlations between size and performance,[13–16]

---

* Associate Director for Research, Office of Research and Development, Indian Health Service, DHHS and Associate Professor (Adjunct), Department of Family and Community Medicine, University of Arizona.

† Senior Scientist, Community Systems Foundation and Professor (Adjunct), Department of Family and Community Medicine, University of Arizona.

‡ Associate Professor, Department of Systems and Industrial Engineering, University of Arizona.

§ Chief, Office of Planning and Development, Navajo Area, Indian Health Service, DHS.

All opinions expressed herein are the authors' and do not necessarily represent the policy of either the Indian Health Service or the Department of Health and Human Services.

Address for reprints: Paul A. Nutting, M.D., Office of Research and Development, Indian Health Service, P.O. Box 11340, Tucson, AZ 85734.

0025-7079/82/0700/0676/$01.25 © J. B. Lippincott Co.

and nearly all have concluded that the relationship is seriously confounded by other factors. Some investigators have concluded that firm performance is independent of structure and that firms can attain their goals by any of a wide variety of structural arrangements.[17] Studies of the relation of size to performance in community agencies have concluded that agencies of different sizes do better at different tasks.[18,19] There is some evidence suggesting that outcomes are better in larger hospitals.[20]

Organizational complexity is another important aspect of structure that has been studied. Weber's[21] pioneering work on bureaucracy stimulated the as yet unsettled debate on the factors needed to describe organizational structure adequately, with complexity playing a central role. Some investigators have found that several dimensions are necessary to describe structure,[17,22] with either size[19,20] or complexity[25] being the more important dimension, while others have found that all structural factors are so highly correlated as to be inseparable.[26–30] Two large studies of government offices found that the logarithm of size was significantly correlated with indicators of complexity, such as horizontal and vertical differentiation, division of labor and personnel specialization.[29–31]

While several studies in the health field vaguely suggest that increased differentiation and specialization may detract from performance,[32–34] other studies of hospitals[35] and industrial firms[36] indicate that complexity interacts with both the coordination and the difficulty of the problems faced by the organization in determining performance. Several studies suggest that greater system complexity *from the viewpoint of the consumer* may detract from performance for high-risk consumers.[9,37]

The study reported in this article seeks to determine how size and organizational type are related to system performance in 11 medical care systems. Performance is defined as the proportion of community members in need of care who receive adequate care according to explicit standards for the process of care. Initially, the study included six Indian Health Service (IHS) units of different size. However, the study was expanded to include three private practices and two HMOs located near the service units, in order to examine performance in three different organizational types, each type with a different payment mechanism. Thus, the study uses cross-sectional data to examine the combined effect of size and organizational type on system performance, within the limitations presented by the 11 study sites.

Study Sites

The 11 study sites include three fee-for-service private practices, two prepaid HMOs and six government-funded IHS service units. First, the six service units were selected from various locations in the western United States to represent the spectrum of service configurations in the IHS. Then, the private practices and HMOs were selected from the general geographic areas of one or more of the study service units. Specifically, private practices B and C are from the same general geographic area as IHS units C and D, while private practice A, both HMOs, and IHS units A and B are near one another.

The precise definition of the "medical care system" at each site presented problems. The "recipient population" was conveniently defined to be the prepaid subscribers at the HMOs; in the private practices, however, in which the recipient population included all past patients potentially still active in the judgment of the private practice, the definition may be inconsistent, and in the IHS units, in which the recipient population included all Native Americans known to be living within the official service area for that site, the population is not always known. In addition, the set of facilities providing service was not well delineated in any of the 11 sites. The study attempted to identify all major sources of medical service for the recipient population of each site, including major referral and "leakage" patterns. Consequently, the "system" as defined for the purposes of the study included facilities and provider groups that did not necessarily consider themselves to be components of a "system," even though patients interacted with the different components as though they were part of the system.

In no instance did a study site that was approached for inclusion in the study decline. In all cases the clinical personnel were cooperative and indicated an interest in using the study results constructively to improve their own organizations.

The typical service unit is located within a rural Native American community and serves a population of 3,000–8,000 people who may reside in smaller communities, often scattered over several thousand square miles. The typical service unit consists of a 30–50 bed hospital and outpatient department staffed by the full complement of health professionals and administrative staff. The main service unit facility serves as the primary referral center and base of operations for one or more full- or part-time field clinics, public health nurses, environmental engineers, health educators and an increasing number of problem-specific health programs (nutrition, mental health, maternal and child health, alcoholism, etc.) operated by the tribal government. Each service unit can refer patients to secondary and tertiary care centers, either operated by the IHS or through contract within the private sector.

However, two of the service units deviate substantially from this pattern. Service unit A is located on the outskirts of a large metropolitan area and consists of a single outpatient clinic staffed by a single general practitioner, a physician extender and a public health nurse. Inpatient care is provided through contractual arrangements with several large hospitals within the metropolitan area. Service unit C is located within another large metropolitan area and consists of a 250-bed medical center, a multidisciplinary outpatient clinic and two field clinics located in the surrounding rural areas. The medical center provides primary care to Native Americans living in the urban area and also serves as a referral center for several other service units, including service unit D. Service unit C is staffed by 40 physicians, of whom 32 are specialists, in addition to a physician extender and five public health nurses. The field clinics are staffed by physicians traveling from the medical center on a weekly and biweekly schedule. All of the IHS units provide primary care to Native Americans at no direct cost to the patient, as well as necessary specialty care requiring referral into the private sector.

All three of the private practices are located in rural areas and are in general proximity to one or more of the service units. Private practices B and C were the only providers of medical care in their communities. All operate on a fee-for-service basis, and participate in Medicare and Medicaid for their eligible patients. The practices consisted of two to five physicians, and one of the groups employed a physician extender. All of the private practitioners were generalists, with the exception of a pediatrician and an internist in private practice C. The physicians of all three practices had admitting privileges at small community hospitals, but none held administrative positions at the hospitals during the study time frame.

The two closed-panel HMOs are located in urban areas and in general proximity to two of the service units. One of the HMOs operated three separate clinics, and physicians in both admitted patients to two separate large urban hospitals. Both had more than 15 physicians on the full-time staff, consisting of both generalists and specialists, and both HMOs employed nurse practitioners in ambulatory care.

Many of the important structural characteristics of the 11 study sites are summarized in Table 1.

The Dependent Variable—Performance

The issues involved in defining performance of a medical care system have been extensively discussed by others.[38] A method for measuring performance that is based on explicit standards for the quality of the process of care as received by all potential consumers (not just patients) for common, ambulatory health conditions was used in the present study. This method considers neither outcomes nor costs. Since a full description of the method and its limitations has been published previously,[39] it is presented here only in summary.

A medical care system should perform various functions, including prevention, screening, treatment and follow-up, for a wide variety of health conditions. In this study, several conditions were used to examine the performance of four functions of care: prenatal and infant care was used to examine prevention performance; prenatal anemia, infant anemia and hypertension were used to examine screening performance; and anemia and urinary tract infection (UTI) were used to examine the performance of both treatment and follow-up. Explicit, minimal standards for the process of care were defined for each condition and function, and are summarized in Table 2. These standards were reviewed and agreed to as appropriate by physicians at every site.

At each study site, a sample of episodes occurring within a particular time frame during the period from 1974 through 1976 was selected for each condition, and the medical records of the persons in the sample were audited to determine if the care they received during the episode met the standard. System performance for a particular function and health condition is defined as the proportion of the entire potential recipient population in need of service that received the service according to the standard. Thus, performance is defined in terms of episodes during which an adequate process of care was received.

No single overall measure of performance for a medical care system is defined, although consistent patterns across functions and health conditions may permit overall observations on the performance of a system. Further, logical sequences of functions, such as treatment and then follow-up, can be combined to obtain an overall performance indicator for the sequence for a particular condition. Logical sequences of tasks within a particular function can also be examined, for example, *contact* between the provider and a consumer-in-need, *recognition* by the provider of the need and, finally, *provision* of appropriate care.

A prevention performance index was calculated as the average score on the six prevention tasks, while the performances for the other functions and conditions were not aggregated. General surveillance was considered part of prevention, while screening for a particular problem, such as anemia, was considered to be a separate function. It is important to note that screening does not include recognition of abnormal results. The conditions chosen are common ambulatory problems that reflect the type of situation most frequently faced by consumer and provider.

The prenatal sample used to examine prevention and screening was taken from a master list of pregnant women compiled from birth certificates, delivery room logs, hospital discharge diagnoses and requisitions for prenatal laboratory work. The master list for the infant sample was compiled from birth certificates and delivery room logs. The sample of persons between 40 and 60 years of age used to examine hypertension screening was generated from a casual selection of active medical records at each site. The master lists for UTI and anemia used to examine treatment and follow-up were obtained from laboratory logs and slip files. In order to maintain the focus on common ambulatory problems, the anemia and UTI samples were limited to simple cases by dropping episodes that showed prior complications, as defined in Table 2. Patients thus eliminated from the sample were replaced from the master list.

Sample sizes for prevention, screening and treatment generally ranged from 27 to more than 100, depending on the health condition and site, but with several exceptions.[a] Samples tended to be smaller for the follow-up function, generally ranging from ten to 40.

Every health record known to exist for each person in all of the samples was sought out in order to capture all contacts between the person and any component of the provider system relevant to the episode in question. For the IHS units, this typically included a review of the medical record at the hospital and at one or more field clinics, public health nursing records and the records at the most frequently used referral centers. For the HMOs and private practices, records reviewed included the office records, hospital medical records and medical records at the most likely sources of referral care. In addition, an attempt was made in each study site to estimate the magnitude of "leakage" of patients outside the study site. When a major pattern was presumed to exist, the records at the major alternate sources of medical care were also audited.

This method of measuring system performance has some properties that avoid certain problems sometimes present in other methods: it samples from the entire community, not just from patients; it examines common ambulatory conditions, not just complex problems requiring hospitalization; it includes and integrates the contributions of many provider disciplines and the consumer; it examines care at many locations; and it examines the performance of several functions of care. Nevertheless, the method also has some important limitations. These include the well-known limitations of information taken from medical records, which may reflect inconsistently actual care in different patients and different systems, and the inconsistent results obtained from different measurements of quality, in which explicit process measures come out lower than other measures. While every effort was made to obtain a representative sample of all community members in need of care, the differing nature of the service populations and the aforementioned difficulty in defining service populations that were comparable across sites may have biased the results. Also, the relative success in identifying all locations at which care was received may have differed across sites. The sensitivity of the standards of care is not well understood, and small changes in the standards may produce significant changes in performance that could alter the findings. The interdependence of the functions is not well understood; for example, screening performance may influence treatment performance. More important, clear relationships between the quality of the process and the outcome have not been established. Thus, there is as yet no guarantee that better performance of process will lead to better outcomes.

The Independent Variables

Two independent variables were used in the analysis: 1) the log to the base 10 of the number of direct care providers (physicians, nurses and physician extenders), which is termed "size"; and 2) a categoric variable termed "type" that reflects whether the system is a private practice, an HMO or an IHS unit. Size is an organizational characteristic that other investigators have often found to be highly correlated with organizational complexity.[29-31] That relation appears to hold in data from this study as well, for example, in the perfect ordinal relation between size and number of disciplines, as shown in Table 1. Use of the logarithm follows previous findings and usage[29-31] and reduces the effect of outlyers.

---

[a] Small samples included infant prevention in HMO-B with five; infant anemia screening in HMO-B with 15; anemia and UTI treatment in IHS-A with 11 and 19, respectively; and anemia treatment in private practices A, B and C with five, nine and 18, respectively. For follow-up the anemia samples were less than ten in private practices A and B, with three and six, respectively.

The most apparent difference between the three organizational types is their payment mechanisms. The private practices are all fee-for-service private entrepreneurships in the traditional style of American medicine. The HMOs have prepaid subscriber plans, with a large portion of the prepayments coming from employers but with subscribers making token payments for many services. The IHS units are government owned and operated, with all provider personnel salaried or under contract. Health care is free to all eligible consumers. There are other differences among the three organizational types that might explain performance variation. Notably, the populations of American Indians and Alaskan Natives served by the IHS units are culturally distinct from the populations served by the HMOs and private practices, which were typical of the general U.S. population.

While the six IHS units vary in size from three to 46 direct care providers, the three private practices are relatively small, ranging from two to five direct care providers, and the two HMOs with 18 and 25 direct care providers are larger than all but one of the IHS units. This clustering of private practices and of HMOs on the size dimension increases the degree to which the two variables confound one another's effects on performance.

Results

Performance scores for the 11 study sites are summarized in Table 3 by function and health condition.# From these data it can be seen that, with the exception of screening for infant anemia, there is wide variation among the study sites for all functions and conditions. IHS service unit A, which is a small system with a largely suburban population, scores consistently higher than the other service units. This site had a stable and, in the authors' opinion, very able team of professionals operating the clinic, which may explain some or all of the high performance. The relative performances of the other ten systems appear to depend on the function and condition, with different systems scoring higher in one area and lower in another.

The primary goal of the analysis was to determine how size and type were related to system performance in the data. The analysis was complicated by the small sample (n = 11) and the confounding of size and type in the data. Analysis of covariance was selected as the best way to structure the analysis and to identify gross relationships, although it was clear that the data do not meet all the assumptions (particularly homoscedasticity) of the statistical model for analysis of covariance.

The combined effect of size and type on six different indicators of performance was obtained using analysis of covariance (Table 4). The null hypothesis in the test for the effect of type was, "After accounting for the effect of size, type had no effect on the variation in performance among the 11 study sites for each function and condition." This hypothesis was rejected only in

---

# The complete data, as well as details of the analysis, are available in: Nutting PA, Burkhalter BR. The Relation of Size, Organizational Type, Tracer, and Sequential Event to Performance in Prevention, Screening, Treatment and Follow-up in Eleven Medical Care Systems: A Working Paper, January 1981. DHHS, IHS, Office of Research and Development, Research and Quality Assurance Branch, Tucson, Arizona

TABLE 1. Structural Characteristics of the 11 Study Sites

Pilot Site	Approximate Population of Catchment Area	Approximate Size of Catchment Area (Sq. mi.)	Fixed Facilities	Number of Direct Care Providers				Number of Community Health Workers	Number of Direct Care[1] Disciplines
				Physicians	Physician Extenders	Public Health Nurses	Total		
IHS A	900	111	1 clinic	1	1	1	3	4	2
IHS B	6,155	2,854	50-bed hospital and OPD; 1 field clinic	5	3	5	13	10	3
IHS C	14,480	92,000	170-bed medical center and multidisciplinary OPD; 2 field clinics	40	1	5	46	35	8
IHS D	3,800	38,000	29-bed hospital and OPD	2	0	3	5	30	2
IHS E	4,926	6,375	41-bed hospital and OPD	5	1	1	7	25	3
IHS F	4,554	5,200	39-bed hospital and OPD; 1 field clinic	4	2	2	8	16	3
Private practice A	20,000	1,600	1 medical office; attending at 30-bed hospital	3	1	0	4	0	2
Private practice B	5,000	1,800	1 medical office; attending at 30-bed hospital	2	0	0	2	0	1
Private practice C	15,000	5,000	1 medical office; attending at 30-bed hospital	5	0	0	5	0	3
HMO A	22,600	2,700	2 medical offices; attending at 250-bed and 300-bed hospitals	20	5	0	25	0	6
HMO B	11,000	1,500	1 medical office; attending at two 300-bed hospitals	16	2	0	18	0	6

The number of disciplines is identified from the following predetermined list: public health nurse, physician-extender, general practitioner, internist, surgeon, pediatrician, gynecologist–obstetrician, super specialist.

TABLE 2. Standards Used to Judge the Adequacy of Care Received and Reported in the Medical Record

Function	Task or Condition	Standard
Prevention	Prenatal diet counseling	Pregnant women should receive diet counseling by 26th week of gestation
	Prenatal surveilance	Pregnant women should have fundal height measured 3 times in 2nd and 5 times in 3rd trimester, and FHR measured once in 2nd and 5 times in 3rd trimester
	Infant immunization	Infants should receive 3 DPT and 2 OPV immunizations by 13 months
	Infant diet counseling	Mothers should receive infant diet counseling 3 times before the infant is 13 months old
	Infant care counseling	Mothers should receive general infant care counseling once in 1st 6 months and once in 2nd 7 months of infant's life
	Infant growth surveillance	Infants should have weight and height measured 3 times in 1st 6 months and 2 times in 2nd 7 months of life
Screening	Prenatal anemia	Pregnant women should have a hematocrit or hemoglobin by 20th week of gestation
	Infant anemia	Infants should have a hematocrit or hemoglobin between 6 and 13 months
	Hypertension	Persons between 40 and 60 years of age should have their blood pressure measured once every 3 years (in the absence of trauma, pregnancy, intoxication or medication known to elevate blood pressure)
Treatment	Anemia	Persons screened positive for anemia (Hct < 33 or Hgb < 11, in the absence of neoplastic disease, acute or chronic blood loss, or anemia previously diagnosed as other than iron deficiency) should contact provider within 3 weeks and start iron therapy within 1 week of contact
	UTI	Persons with positive urine culture (more than 100,000 organisms, in the absence of chronic or recurrent UTI or known abnormal anatomy of urinary tract) should start appropriate antibiotic therapy within 2 weeks
Follow-up	Anemia	Persons on iron therapy for anemia should have a hematocrit or hemoglobin between 3 and 6 weeks after iron therapy was started
	UTI	Persons on antibiotic therapy for UTI should receive a urine culture or microscopic urinalysis between 1 and 4 weeks after antibiotic therapy was started

screening for infant anemia, where the significance level is 0.004.

The null hypothesis for the effect of size was, "After acconting for the effect of type, size had no effect on the variation in performance among the 11 study sites for each function and condition." This hypothesis was rejected in the treatment and follow-up sequence for both anemia and UTI, and was nearly rejected in screening for hypertension. The relationship between size and performance is strongly negative in these cases, with the smaller systems performing better than the larger ones.

These results from the analysis of covariance must be viewed as indictors of gross

TABLE 3. Performance in 11 Medical Care Systems by Functions, Health Condition, Type of Organization and Size

Function and Condition	PP			IHS						HMO		AVERAGE			
	B	A	C	A	D	E	F	B	C	B	A	PP	IHS	HMO	ALL
Prevention															
Six task average	5	19	17	56	20	21	7	26	28	26*	47	14	26	36	25
Screening															
Prenatal Anemia	49	42	51	56	48	34	33	36	43	89	71	47	42	80	50
Infant Anemia	39	30	39	36	34	31	33	32	36	33*	31	36	34	32	34
Hypertension	86	86	79	92	84	42	34	62	41	64	62	84	59	63	67
Treatment and follow-up															
Anemia	33*	40*	39*	36*	36	18	10	14	6	18	11	37	20	15	24
UTI	65	29	42	47*	43	22	16	20	8	26	18	45	26	22	31
Number of direct care providers	2	4	5	3	5	7	8	13	46	18	25				

The performance scores reflect the percentage of consumers in need of care who received adequate care for the condition and function indicated. Treatment and follow-up is reported here as a single function, with the performance scores reflecting the percentage of consumers in need of treatment who received both adequate treatment *and* adequate follow-up care. Asterisks (*) indicate a small sample (less than 25). The three organizational types are private practice (PP), Indian Health Service units (IHS) and health maintenance organizations (HMO).

relationships rather than as reliable statistical conclusions, for the reasons noted. From this viewpoint, the strong inverse relationship of size with some of the performance indicators really reflects the existence of an inverse relationship of size and performance in the data of six IHS units, with the three small private practices and two large HMOs more or less fitting into the service unit pattern.

Prevention

While neither size nor type was significantly related to prevention performance in the analysis of covariance, the *average* prevention performance of the two HMOs was considerably higher than the average performance of the three private practices or of the six IHS units. Several outlyers with especially high scores (HMO-A and IHS-A) or especially low scores (PP-B and IHS-F) contributed to this situation by increasing the variance of performance and confounding the effect of size and type even more. The differences among the three types are not consistent across the six prevention tasks. For example, the private practices scored highest in prenatal surveillance but lowest in the infant care tasks, while the IHS service units averaged highest in infant immunization and infant care counseling. Several study sites had one or more zero or near-zero performance scores, which pulled down their overall prevention performance, suggesting that some tasks were simply ignored by some of the provider systems. In fact, one service unit (IHS-C) noted that it was not performing some of the prevention tasks at the standards review and indicated its intention to change. Although the large average differences between types in prevention performance may justify further examination of the prevention function with a larger sample of medical care systems, our data certainly do not support the existence of a significant relationship.

TABLE 4. Results of an Analysis of Covariance to Find the Relation of System Type and Size to System Performance in 11 Medical Care Systems

Function and Condition	Significance of Type	Significance of Size	Direction of Relation of Size to Performance
Prevention			
Six task average	0.385	0.727	(0)
Screening			
Prenatal anemia	0.004	0.364	(−)
Infant anemia	0.648	0.942	(0)
Hypertension	0.435	0.077	(−)
Treatment and follow-up			
Anemia	0.470	0.028	(−)
UTI	0.511	0.009	(−)

Type is a categoric variable reflecting whether the system is a private practice, HMO or IHS unit. Size is the logarithm of the number of direct care providers. Performance is the proportion of consumers receiving adequate care according to explicit standards by function and condition. Treatment and follow-up is reported here as a single function. Each significance level is the probability that the variance in the performance data explained by the independent variable in question is a chance occurrence after first accounting for the variance explained by the other independent variable. The analysis of covariance used was: Subprogram ANOVA, option 7 (in Nie NH, et al. SPSS, 2nd edition. New York, McGraw-Hill, 1975). The results reported here are not necessarily statistically valid because the assumptions of the covariance model are not met, particularly homoscedasticity, and, in addition, the two independent variables (Type, Size) are seriously confounded.

Screening

Overall, approximately half of the consumers-in-need received adequate screening, but the performance differed significantly ($p < 0.001$) by health condition, with screening performance for infant anemia (34 per cent) being considerably below the performance for either prenatal anemia (50 per cent) or hypertension (67 per cent). Table 3 shows that the average screening performances are higher than either prevention performances (which were held down by numerous zero scores) or the treatment/follow-up performance. The latter reflects the longer sequence of tasks that must be performed correctly in the treatment/follow-up function and therefore allows greater possibility of error. In fact, treatment performances alone are generally higher than screening performances.

The effect of size and type on screening performance also differed by health condition. For infant anemia, neither size nor type was important; for prenatal anemia, type was significantly related to performance (at the 0.004 level), but size was not; and for hypertension, size was related to performance (at the 0.077 level), but type was not. For prenatal anemia, the average performance of the HMOs (80 per cent) was considerably higher than the average performance of either the private practices (47 per cent) or the IHS units (42 per cent).

Screening requires the successful completion of two events in sequence—the first is *contact* between the consumer and provider, and the second is *provision* of an appropriate screening test once contact has been made. Provision of screening once contact had been made was performed adequately in about the same proportion of consumers-in-need in all three health conditions (averaging 66 per cent in infant anemia, 74 per cent in prenatal anemia and 75 per cent in hypertension), while contact rates differed considerably among the health conditions (averaging 53 per cent in infant anemia, 74 per cent in prenatal anemia and 87 per cent in hypertension).

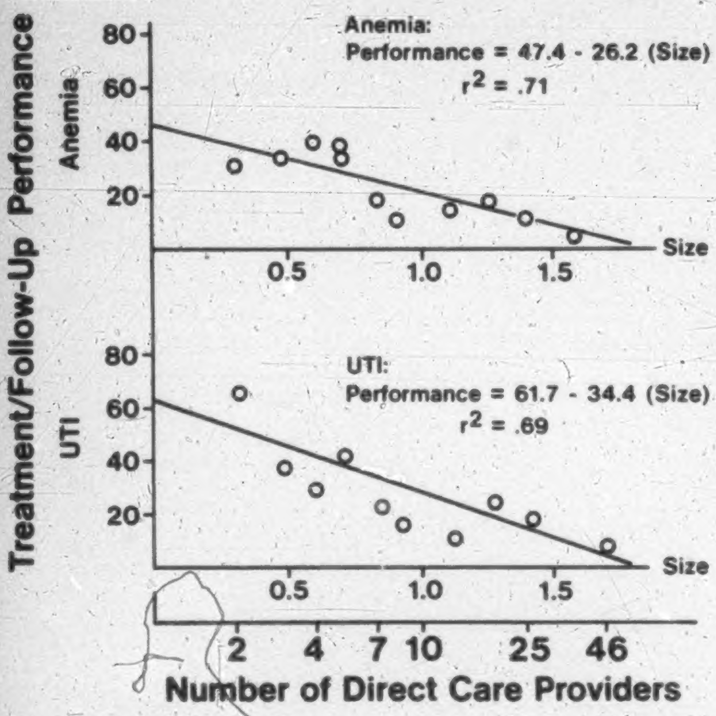


FIG. 1. Linear regression of size with treatment /follow-up performance for anemia and UTI in 11 medical care systems, where size is the logarithm of the number of direct care providers in the system and performance is the proportion (out of 100) of consumers in need of care who received adequate treatment and follow-up care. This figure demonstrates the increasing performance in curative care in smaller systems.

Contact rather than provision was responsible for the superior performance of the HMOs in relation to the other types in prenatal anemia screening. However, in hypertension screening, the inverse relation of size to performance appears primarily in provision rather than contact.

Treatment and Follow-up

Treatment and follow-up have been combined into a single sequence of care and one measure of performance, used for both functions, that reflects the proportion of consumers in need of treatment who received both adequate treatment and adequate follow-up care. This was done 1) because treatment and follow-up are a logical sequence, and the same sample of episodes was used to explore both functions in each study site, and 2) because the sample sizes for follow-up are rather small and biased in favor of individuals who receive adequate treatment in that they contain the subset of the treatment sample that received adequate treatment.

The results are dramatic. Size shows a strong, significant and negative correlation with performance of treatment/follow-up while type is not related. Small systems do better, as seen in Tables 3 and 4, and in Figure 1. The analysis of covariance produced significance levels for the effect of size on performance, after accounting for type, of 0.028 for anemia and 0.009 for UTI, while both regression lines in Figure 1 differ from zero at a significance level less than 0.005.

The inverse relationship of size to treatment/follow-up performance is not only statistically significant; it is also large. The smaller systems (2–5 direct providers) experienced a 30–50 per cent performance in treatment/follow-up, compared with 10–25 per cent in the larger systems (seven or more direct care providers); in other words, the smaller systems were roughly twice as successful as the larger ones. Within the range of the present data, when the number of direct care providers doubled, the performance was reduced by approximately another ten percentage points.

The treatment and follow-up function requires successful completion of six events in sequence: *contact* between the consumer and provider for treatment; *recognition* by the provider that treatment is needed; and then *provision* of the needed treatment, followed by contact, recognition and provision of follow-up care. In order to gain more insight into the strong and significant inverse relation between size and the performance of the treatment/follow-up function, the performance of various components of treatment and follow-up was examined, including the relation of size to treatment only; to follow-up only; and to contact, recognition and provision of service. While average treatment performance (72 per cent) was superior to follow-up performance (41 per cent), as shown in Table 5, the significant inverse relation between size and performance held for both treatment and follow-up. When size was regressed with performance of anemia treatment, with performance of UTI treatment, with performance of anemia follow-up and with performance of UTI follow-up, the slope was negative and differed from zero at the 0.05 significance level in all four cases.

In addition, size was regressed against all contact performances (anemia treatment contact, UTI treatment contact, anemia follow-up contact and UTI follow-up contact) for each of the 11 study sites, and also with all recognition and provision performance scores. The results are shown in Table 5. Average provision performance (88 per cent) is superior to the average performance of either contact (77 per cent) or recognition (80 per cent). The regression lines are negative in slope for all three sequential events, but significantly so only for recognition and provision. IHS units with larger catchment areas have more portals of entry in the form of more field clinics and community health workers than the other study sites, a strategy that

TABLE 5. The Results of Regression Analysis Relating Size to Performance of Contact, Recognition, and Provision in Treatment and Follow-up for Anemia and UTI, and Comparing Average Treatment and Follow-up Performances

Regression Equation	r^2	Probability that Slope = 0	Average Treatment Performance	Average Follow-up Performance	Overall Average Performance
Contact performance = 83.9 − 8.1 (Size)	0.05	p = large	85%	68%	77%
Recognition performance = 97.0–24.4 (Size)	0.32	p = 0.001	84%	66%	80%
Provision performance = 100.6–13.6 (Size)	0.31	p = 0.001	90%	87%	88%
Percent of consumers with adequate contact, recognition and provision			72%	41%	

This table demonstrates the negative and, for recognition and provision, statistically significant relationship between size and performance of the sequential events of contact, recognition and provision in the treatment and follow-up functions for anemia and UTI. Treatment performance is superior to follow-up performance. Each regression equation used 44 observations, with each regression using the following four observations at each of the 11 sites: anemia treatment, anemia follow-up, UTI treatment and UTI follow-up.

may be responsible for maintaining their contact performance roughly equal to that of the geographically smaller systems.

Thus, the inverse relation between size and performance during treatment and follow-up is pervasive in these data, being found in both anemia and UTI, in both treatment and follow-up independently, and in recognition and provision, although the samples used to examine follow-up as an independent function are small and biased, as noted.

Discussion

Eleven medical care systems of varying size and payment mechanisms (consisting of six government funded and operated IHS units, three fee-for-service private practices and two prepaid HMOs) were analyzed with respect to their performance for various functions (prevention, screening, treatment and follow-up) and health conditions (prenatal and infant care, hypertension, anemia and UTI). Size was defined as the logarithm of the number of direct care providers in each system, while performance was the percentage of potential consumers in need of care who received adequate care according to explicit standards for the process of care.

Size was found to have a strong, significant and negative correlation with performance of both the treatment and follow-up functions, but payment mechanism was not. Within treatment and follow-up the strong inverse relation was also true for recognition and provision, but not for contact. In screening, the relationship of size and payment mechanism to performance depended on the health condition: prepaid HMOs performed significantly better in screening for prenatal anemia, but size was not a significant factor; small systems performed better than large systems in screening for hypertension, but payment mechanism was not a significant factor; neither size nor payment mechanism influenced screening for infant anemia. In prevention, neither size nor payment mechanism was statistically important in explaining performance, although the HMOs had a higher average prevention performance than the service units or private practices. There was no evidence to contradict the previously reported finding[10] that HMOs provide better preventive care, because the marginal cost of preventive care to the consumer is very low.

These results tend to support previous findings that organizations of different size do relatively better at different tasks.[18,19] In the present case, smaller organizations tended to do relatively better at curing tasks than at screening or prevention. Smaller systems were also better at recognition, which fits previous findings that continuity of provider increases recognition performance[34] if one assumes that patients are less likely to encounter the same provider in a larger system. The higher provision performance observed in smaller systems may also be due to greater continuity of provider. In larger systems, there may be not only a greater liklihood that patients will contact a different provider in a different location from one visit to the next, but also a greater liklihood that the provider who is contacted will not be able to perform the needed service, even though he/she recognizes the need for it. For example, a public health nurse is not trained to prescribe certain treatments, and a field clinic may not have the facilities for certain procedures. Of course, larger systems can organize themselves in various ways to try to maximize continuity of provider and minimize wasted encounters, although systematic efforts by the larger systems in this study to do that were not evident. The fact that contact performance did not vary significantly with size suggests that the multiple portals of entry in the large IHS systems have been effective in increasing access to care.

The data from this study also tend to support previous findings of high correlations between size and complexity, although the finding in this regard is based on very little data. The data also support previous findings[29–31] that the logarithm of size relates more strongly to complexity than does size alone.

With the exception of screening for prenatal anemia, organizational type does not appear to be a very important predictor of performance, and, thus, the differences in payment mechanisms represented here do not appear to be important.

However, these conclusions must be viewed within the limitations of the study, and certain caveats must be considered before generalizing the results to other settings. First, although there was no apparent bias due to self-selection (no site approached for the study declined participation), the study sites were not chosen randomly and are not necessarily a representative cross section of their organizational types. Second, no urban or large multispecialty private practices were included in the study, and no inference should be made that the three rural private practices represent the much larger universe of fee-for-service medical care. Third, the number of sites studied was small (only 11), and in some cases the number of observations for particular functions and health conditions were also quite small. The confounding of size with type of payment mechanism is a further limitation resulting partially from the small sample (for example, the sample contained only small private practices and only relatively large HMOs). Fourth, the study examined the performance of the medical care system only for basic issues in medical care. The strongly negative correlations observed in treatment and follow-up might not have been present if the study had focused on more complex medical problems. Fifth, the performance measure itself has a number of limitations, as summarized above. Sixth, numerous other variables that have not been systematically considered here might better explain the observed variation in performance. One of these, management capability, was observed informally and, except for the case noted above, did not appear to vary noticeably across study sites. Other aspects of organizational complexity, such as the relative use of physician extenders and public health nurses in place of physicians, may also be important. Finally, of course, this is a study of correlations among cross-sectional data, with no claim to making causal inferences based on appropriate controls.

In light of these limitations, the results should be the cause for hypotheses rather than the cause for action. Do the results hold for a larger and better controlled sample of medical care systems? Do they hold for a wider variety of health conditions? Does the apparent inability of the larger systems in this study to provide adequate treatment and follow-up care reflect simply a lack of management technology for control and coordination, as suggested by some investigators,[38,39] or does it reflect faults in basic assumptions underlying the medical care delivery system, as suggested by others?[40–42] Can the suspected deleterious effects of organizational size be corrected by organizational arrangements, such as subgroups? Would the introduction of costs into the definition of performance alter the results? In any case, the results of this study may have important implications for those who advocate increasingly large and complex health care systems to meet the growing demands for the quality and quantity of medical services.

References

1. Pauly MV. Efficiency, incentives, and reimbursement for health care. Inquiry 1970;7:114.

2. Dutton DB. Patterns of ambulatory health care in five different delivery systems. Med Care 1979;17:221.

3. Roemer M, Shonick W. HMO performance: the recent experience. Milbank Mem Fund Q 1973;51:271.

4. Slesinger D, Tessler R, Mechanic D. The effects of social characteristics on the utilization of preventive medical services in contrasting health care programs. Med Care 1976;14:392.

5. Gaus DR, Cooper BS, Hirschman CG. Contrasts in HMO and FFS performance. Soc Security Bull 1976;39:5.

6. LoGerfo JP, Larson E, Richardson WC. Assessing the quality of care for urinary tract infection in office practice: a comparative organizational study. Med Care 1978;16:488.

7. Shortell S, Richardson W, LoGerfo JP, et al. The relationships among dimensions of health services in two provider systems: a causal model approach. J Health Soc Behav 1977;18:139.

8. Richardson WC, Diehr PK, LoGerfo JP, et al. Comparisons of prepaid health care plans in a competitive market: the Seattle prepaid health care project. DHEW Publ No (PHS) 80-3199. NCHSR Research Summary Series, August 1980.

9. Wolinsky FD. The performance of health maintenance organizations: an analytic review. Milbank Mem Fund Q 1980;58:537.

10. Luft HS. Why do HMOs seem to provide more health maintenance services? Milbank Mem Fund Q 1978;56:140.

11. Miller EM. The extent of economies of scale—the effects of firm size on labor productivity and wage rates. Southern Econ J 1978;44:470.

12. Rakowski JP. Cost differences according to firm size in U.S. trucking. Transportation J 1980;18(2):63.

13. Singh A, Whittington G. Growth, Profitability and Valuation. Cambridge: Cambridge University Press, 1968.

14. Bethlehem G. Note on growth, profitability and size of firms. S Afr J Econ 1979;47:198.

15. Soete LLG. Firm size and inventive activity—evidence reconsidered. Eur Econ Rev 1979;12:319.

16. Buckley PJ. Influence of firm size, industry, nationality, and degree of multinationality on growth and profitability of world's largest firms. Weltwir Arc 1978;114:243.

17. Reimann BC. On the dimensions of bureaucratic structure: an empirical reappraisal. Adm Sci Q 1973;18:462.

18. Ostrom E, Parks RB, Whitaker GP. Some evidence of effects of policy agency size. The Police Chief 1978;45(2):28.

19. Kovack KA. Organization Size, Job Satisfaction, Absenteeism and Turnover. Washington, D.C.: University Press of America, 1977. (Reviewed by Muchinsky PM in Personnel Psych 1978;31:436.)

20. Scott Wr, Flood AB, Ewy W. Organizational determinants of services, quality and cost of care in hospitals. Milbank Mem Fund Q, 1979;57.

21. Weber M. The Theory of Social and Economic Organization. Translated by Henderson AM. Parson T. New York: The Free Press, 1946.

22. Pugh DS, Hickson DJ, Hingings CR, Turner C. Dimensions of organization structure. Adm Sci Q 1968;13:65.

23. Shortell SM, Getzen TE. Measuring hospital medical staff organizational structure. Health Sys Res 1979;14(2):97.

24. Terrien FC, Mills DC. The effect of changing size upon the internal structure of an organization. Am Sociol Rev 1955;20:11.

25. Anderson TR, Warkov S. Organizational size and functional complexity. Am Sociol Rev 1961;26:23.

26. Hall RK. Intraorganizational structural variation: application of the bureaucratic model. Adm Sci Q 1962;7:295.

27. Blau PM. The hierarchy of authority in organizations. Am J Sociol 1967;23:453.

28. Child J. Organization structure and strategies of control: a replication of the Aston studies. Adm Sci Q 1972;17:163.

29. Beyer JM, Trice HM. A reexamination of the relations between size and various components of organizational complexity. Adm Sci Q 1979;240:48.

30. Blau PM. A formal theory of differentiation in organizations. Am Sociol Rev 1970;35:201.

31. Blau PM, Schoenherr RA. The Structure of Organizations. New York: Basic Books, 1971.

32. Morehead MA, Donaldson R. Quality of clinical management of disease in comprehensive neighborhood health centers. Med Care 1974;12:302.

33. Morehead MA. The medical audit as an operational tool. Am J Public Health 1967;57:1643.

34. Starfield BH, Simborg DW, Horn SD, Yourtee SA. Continuity and coordination in primary care: their achievement and utility. Med Care 1976;14:625.

35. Longest B. Relationships between coordination, efficiency and quality of care in general hospital. Hosp Adm 1974;19(14):65.

36. Lawrence PR, Lorsch JW. Differentiation and integration in complex organizations. Adm Sci Q. 1967; 12:1.

37. Nutting PA, Barrick JE, Logue SC. The impact of a maternal and child health care program on the quality of prenatal care: an analysis by risk group. J Community Health 1979;4:267.

38. Donabedian A. Needed Research in the Assessment and Monitoring of the Quality of Medical Care. National Center for Health Services Research Report, July 1978. (DHEW Publication No. (PHS) 78-3219.)

39. Nutting PA, Shorr GI, Burkhalter BR. Assessing the performance of medical care systems: a method and its application. Med Care 1981;19:281.

40. Navarro I. The underdevelopment of health of working America: causes, consequences, and possible solutions. Am J Public Health 1976;66:538.

41. Illich I. Medical Nemesis: The Expropriation of Health. London: Calder and Bayars, 1975.

42. Carlson R. The End of Medicine. New York: John Wiley and Son, 1975.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Effect of a WIC Program on Children's Clinic Activity in a Local Health Department

JONATHAN B. KOTCH, M.D., M.P.H.,* AND DAVID WHITEMAN, M.A., PH.D.†

Two years' encounter data from a local health depratment in rural North Carolina are analyzed to determine the extent to which WIC increased the demand for health department services. WIC clinic visits were found to contribute disproportionately to Children's Clinic activity. Nearly 80 per cent of the increase in total clinic encounters from FY78 to FY79 was attributable to WIC alone. Whereas in the first quarter of FY78 only 18 per cent of Children's Clinic encounters by infants were to WIC clinics, that number had risen to more than 58 per cent by the last quarter of FY79. The number of children on WIC as a percentage of all non-WIC encounters with public health nurses increased during the same period from 19 to 34 per cent. Finally, infants who came for the first time to the Children's Clinic to be certified for WIC contributed an increasing percentage of all infants seen in non-WIC clinics, from 0 per cent at the start of the study to almost 15 per cent by the final quarter. It is concluded that the presence of the WIC program in this local health department was responsible for substantial increases in non-WIC clinical activity that were not supported financially by WIC or any other source.

THE SPECIAL SUPPLEMENTAL FOOD PROGRAM for Women, Infants, and Children (WIC) is now one of the "most effective and successful" child nutrition programs in the country.[1] Throughout the formal consideration of this program, which now spans more than a decade of debate, demonstration and nationwide implementation, its proponents have based their support on the significant long-term health benefits resulting from food supplements for pregnant women, breast-feeding mothers, and infants and children under five years of age who are at nutritional risk. On the strength of this argument, the WIC program has grown from a $20-million pilot study in 1972 to a $750-million program in 1979.

Most evaluation studies of WIC have focused on the physiologic and anthropometric effects of nutrition supplementation on mothers and children.[2,3] The focus of this article, however, is the consequences of the decision to place WIC programs in health agencies. Two arguments have been advanced in support of a health locus for WIC. One argument is that such placement would "encourage low-income families to come to the health centers so that total maternal and child care services can be provided."[4] A second argument can be traced to the urgings of Senator Hubert Humphrey, author of the original legislation, who said that such placement would ensure that the

---

* Department of Maternal and Child Health, School of Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

† Department of Government and International Studies, University of South Carolina, Columbia, South Carolina.

Supported in part by the North Carolina Area Health Education Center (AHEC) Program. Dr. Whiteman was supported by a research assistantship provided by the Department of Maternal and Child Health, School of Public Health, The University of North Carolina at Chapel Hill.

Address for reprints: Jonathan B. Kotch, M.D., M.P.H., Department of Maternal and Child Health, School of Public Health, The University of North Carolina at Chapel Hill, Rosenau Hall 201-H, Chapel Hill, NC 27514.

0025-7079/82/0700/0691/$00.90 © J. B. Lippincott Co.

program would "not be mismanaged in terms of some of the other programs."[5] Recent Congressional debate has made more clear the desire of WIC supporters that the program be perceived as a health, rather than a welfare, program.[6]

However, throughout Congressional consideration of the WIC program, supporters have failed to anticipate the additional demands that WIC would make on already overcrowded health department services, particularly on those provided by public health nurses. Similarly, past evaluators of the WIC program have failed to take this into consideration. This article attempts to address this deficiency by exploring the experience of one public health department in a rural North Carolina county.

Background

The population of Columbus County, in southeastern North Carolina, was 91 per cent rural in 1970. The per capita income, $4,349 in 1977, ranked 87th among the state's 100 counties. The infant mortality rate, 23.3 per 1,000 live births in 1977, is higher than that of the state as a whole, which in turn is higher than that of the nation. There are no pediatricians in the county, and its physician/population ratio, 2.4 per 10,000 in 1975, puts it in the bottom fifth of the state.

In late 1974, the County Health Department opened a Children's Clinic, utilizing the services of its newly trained pediatric nurse practitioner, a public health nurse, an outreach worker and a clerk. In March 1977, Columbus was one of four counties that undertook jointly a WIC program. With the arrival of the county's WIC nutritionist in November 1977, the program's operation began in earnest.

Methods

The conclusions of this study are based largely on data collected through June 30, 1979, the end of the second full year that a computerized, encounter-based information system was in operation in the Children's Clinic.‡ Note that, except for the special analysis described below, the unit of measurement is the encounter, not the child. For the purpose of assessing changes in Clinic activity, the number of encounters is more appropriate than an unduplicated count of children. During these two years, the Children's Clinic recorded a total of 5,983 encounters with infants, children and youths aged 0–21 years.

Not all Health Department encounters with children occur in the Children's Clinic, so the information system does not capture the universe of all child visits. For example, infants and children presenting only for immunization do not go through the Children's Clinic. All child visits for health screening, health maintenance, acute illness and follow-up care occur in the Children's Clinic.

Encounters have been classified into three types: WIC, well baby/well child ("well"), and all other encounters ("other"). The WIC category includes clients who come to the clinic for semi-annual WIC certification or recertification, but not clients who come for WIC nutritional counseling. Such counseling visits with the WIC nutritionist, which do not entail a clinical service, are not included as encounters. The "other" category includes such services as episodic visits, follow-ups laboratory screening only and referrals from the nurse to the physician.

A special subanalysis was undertaken to track individual children who first appear at a WIC clinic but who later visit a non-WIC clinic. In order to capture that population of children whose first visit to any Health Department clinic occurred during the period of study, only children born on or after July 1, 1977, are included in the subanalysis. Of the total of 5,983 Children's Clinic encounters recorded between July 1, 1977, and June 30, 1979, 2,946 were accounted for by this group of 1,258 infants.

---

‡ Since September 1976, activity in the Columbus County Children's Clinic has been monitored by the Maternal and Child Health Department of the University of North Carolina School of Public Health. The Department has also recruited physicians who, while Master's candidates in the school, provide pediatric backup for the clinic's nurses. After several months of trial operation, a computerized information system was fully implemented on July 1, 1977. The system provides information of every visit to the Children's Clinic, including encounter type, provider type, health problem(s), referral(s), screening test result(s), and demographic and anthropometric data.

Results

The assessment of WIC impact on Children's Clinic activity can be separated into two parts. First, the general impact of the program on total clinic activity can be ascertained by examining the frequency of all clinic encounters, including WIC visits. Second, the more specific impact of WIC on the activities of public health personnel can be analyzed by examining the frequency of encounters processed by the Clinic's nursing staff.

In the first case, total activity in the Clinic increased substantially during the two-year period. Total Clinic encounters increased 32.5 per cent from FY78 to FY79. Figure 1 displays quarterly data divided into the three encounter categories. The rate of increase for WIC encounters is much greater than the rate for the other two categories. According to Table 1, while the "well" encounters increased by 19 per cent and the "other" encounters increased by five per cent, the number of WIC encounters grew from 1,241 to 1,910, an increase of 54 per cent. The nursing staff saw 2,240, or 71 per cent, of these 3,151 encounters, 926 in FY78 and 1,314 in FY79.

Table 1 also reveals some seasonal oscillation in Clinic activity. By comparing the total encounters for each quarter of the second year with those of the same quarter of the first year, such seasonal variation can be controlled. The resulting figures indicate that the increase in WIC encounters in all cases made the largest single contribution to the total increase between each pair of quarters. "Well" and "other" encounters also contributed to the overall increase, but to a lesser extent.

In order to examine further the demand that WIC clients made on activity in well-child and other clinics (i.e., "non-WIC" clinics), a subanalysis of infants born on or after the initiation of data collection was undertaken. This analysis shows that the proportion of all visits made by this group to a WIC clinic increased from 37.7 per cent in the first fiscal year to 62 per cent in the second. Similarly, when one looks only at first visits made by these infants, 36.4 per

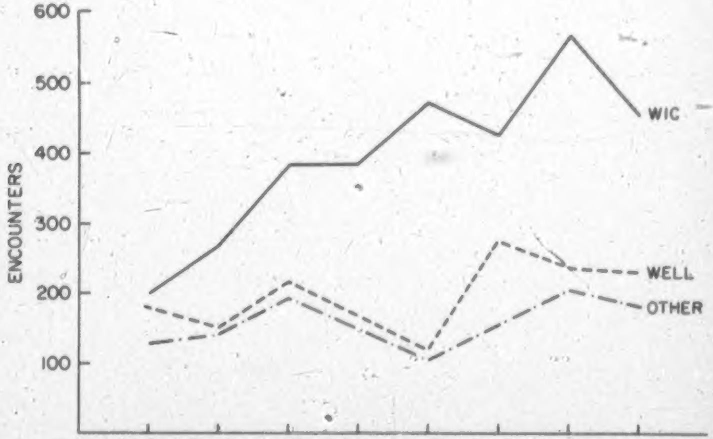


FIG. 1. Total clinic encounters by type of encounter, Columbus County Children's Clinic, July 1, 1977, through June 30, 1977 by quarter year.

TABLE 1. All Encounters per Quarter and per Year, by Encounter Type, Columbus County Children's Clinic, FY78 and FY79

	Quarter Years				Total		
Encounter Type	July–Sept.	Oct.–Dec.	Jan.–March	April–June	Annual Number	(Per Cent Increase)	Per Cent Contribution to Total Annual Increase
WIC							
FY79	471	425	563	451	1910		
FY78	199	271	385	386	1241		
Increase	272	154	178	65	669	(53.9)	79.9
Well							
FY79	117	273	234	227	851		
FY78	178	150	215	171	714		
Increase	−61	123	19	56	137	(19.2)	16.4
Other							
FY79	104	160	202	183	649		
FY78	128	143	194	153	618		
Increase	−24	17	8	30	31	(5.0)	3.7
Total							
FY79	692	858	999	861	3410		
FY78	505	564	794	710	2573		
Increase	187	294	205	151	837	(32.5)	100

cent in the first year were to a WIC clinic, compared with 58.6 per cent in the second. Table 2 displays the encounter data for these infants for each individual quarter year.

The extent to which those infants who initially came to the Health Department to be certified for WIC but who returned for a non-WIC clinic was addressed by looking at the unduplicated counts of such infants as a proportion of infants visiting all children's clinics and as a proportion of those infants visiting only non-WIC clinics (Fig. 2). In both cases, the proportions increased. In the first quarter, July 1 through September 30, 1977, no infants visited a WIC clinic before visiting a non-WIC clinic. However, the percentage of infants seen in any encounter in the Children's Clinic whose first visit ever was to a WIC clinic rose from 2.10 per cent in the second quarter to 5.39 per cent in the last quarter. An even steeper increase was observed when the unduplicated number of infants visiting non-WIC clinics only is examined. The percentage of such infants who first visited a WIC clinic rose from 3.19 per cent in the second quarter to 14.81 per cent in the final quarter (Table 3).

The influence of the expanding WIC Program on the work load of existing public health personnel can be assessed by examining encounters for all children with the Clinic's two full-time nurses. The total number of all categories of encounters with the nursing staff increased 33.6 per cent during the study period, a slightly greater percentage increase than that for all Clinic providers combined. The addition, in November 1977, of a full-time WIC nutritionist did not mitigate the new demands made on nurses' time.

Nursing staff encounter figures for similar quarters are compared in Table 4. The results differ from the results of all encounters (Table 1) in that, in one pair of matched quarters, the "well" category is the largest component of the total increase in nursing

TABLE 2. Visits to Columbus County WIC Clinics by Infants Born on or After July 1, 1977, by Quarter Year

	All Visits			First Visits		
		WIC Clinics			WIC Clinics	
	All Clinics	Number	Per Cent	All Clinics	Number	Per Cent
1977						
July–Sept.	154	28	18.2	125	27	21.6
Oct.–Dec.	176	54	32.4	102	43	42.2
1978						
Jan.–March	306	136	44.4	160	74	46.3
April–June	306	134	43.8	132	45	34.1
July–Sept.	447	295	66.0	215	148	68.8
Oct.–Dec.	442	280	63.4	175	104	59.4
1979						
Jan.–March	621	380	61.2	195	99	50.3
April–June	494	288	58.3	152	82	54.0

staff activity. However, the WIC increase is the largest component in the other three cases and consequently accounts for the largest proportion of the total annual increase. WIC was also the only category of encounters to show an increase in each of the four quarters, including the first quarter when total encounters went down.

One final indicator of WIC's influence on nursing staff activity, the proportion of

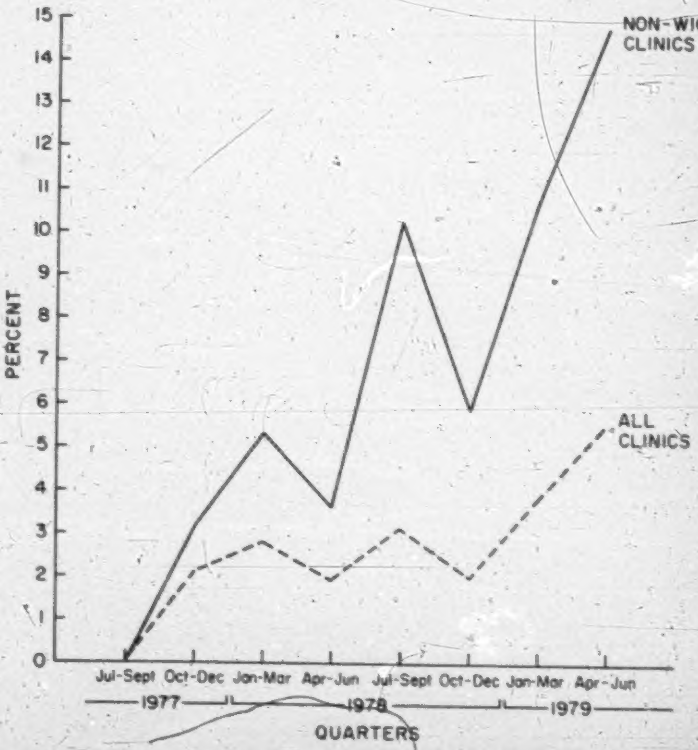


FIG. 2. Unduplicated infants whose first visit to Columbus County Children's Clinic was to a WIC clinic, as a percentage of all infants born on or after July 1, 1977, who were seen in all clinics or in non-WIC clinics, by quarter year.

TABLE 3. Unduplicated Counts of All Infants Born on or After July 1, 1977, and of Such Infants Whose First Visit to Columbus County Children's Clinic Was to a WIC Clinic, Seen in All Clinics and in Non-WIC Clinics, by Quarter Year

	All Clinics			Non-WIC Clinics Only		
		Infants First Seen for WIC			Infants First Seen for WIC	
	Total Infants	Number	Per Cent	Total Infants	Number	Per Cent
1977						
July–Sept.	125	0	0	98	0	0
Oct.–Dec.	143	3	2.10	94	3	3.19
1978						
Jan.–March	251	7	2.79	130	7	5.38
April–June	258	5	1.94	138	5	3.62
July–Sept.	407	13	3.19	127	13	10.24
Oct.–Dec.	395	8	2.03	135	8	5.93
1979						
Jan.–March	526	20	3.80	186	20	10.75
April–June	430	24	5.58	162	24	14.81

"well" and "other" encounters made by clients already on WIC, is shown on Table 5. The data, available only after April 1978, demonstrate a steady increase. During the 1978 quarter from April to June, clients on WIC accounted for 19 per cent of non-WIC encounters. By the final quarter, such clients accounted for 34 per cent of non-

TABLE 4. Nursing Encounters per Quarter and per Year, by Encounter Type, Columbus County Children's Clinic, FY78 and FY79

	Quarter Years				Total		
Encounter Type	July–Sept.	Oct.–Dec.	Jan.–March	April–June	Annual Number	(Per Cent Increase)	Per Cent Contribution to Total Annual Increase
WIC							
FY79	199	304	457	354	1314		
FY78	148	210	313	255	926		
Increase	51	94	144	99	388	(41.9)	62.6
Well							
FY79	87	260	198	198	743		
FY78	159	144	185	124	609		
Increase	−69	116	13	74	134	(22.0)	21.6
Other							
FY79	56	95	135	122	408		
FY78	69	48	104	89	310		
Increase	−13	47	31	33	98	(31.6)	15.8
Total							
FY79	342	659	790	674	2465		
FY78	373	402	602	468	1845		
Increase	−31	257	188	206	620	(33.6)	100

WIC encounters. Therefore, not only did the "well" and "other" nursing encounters increase substantially (22 and 32 per cent, respectively) from one year to the next, but the proportion of encounters attributable to children on WIC seen in those clinics increased steadily.

Discussion

Congress passed the WIC Program with the dual intent of conferring the health benefits of food supplements on pregnant women, breast-feeding mothers, and infants and children at risk, and of improving the overall health status of program participants. The legislative history of the program indicates that, in placing the WIC Program in health agencies, sponsors hoped to avoid the onus of a "welfare program" label. Little serious thought was given to the consequences of attracting a new constituency to local health departments. In particular, no provision was made for the increased demand WIC clients would make on those health department staff also providing non-WIC services.

It has been demonstrated that the total encounters in the Columbus County Children's Clinic increased remarkably during a two-year period, coinciding largely, but not completely, with the growth of the WIC program. There were 32.5 per cent more total encounters in FY79 than in FY78. Total WIC encounters increased 54 per cent during the same period. This is consistent with an increase in the average monthly active case load of WIC infants and children of 52 per cent over the same period.§

The attempt to substantiate the impression that the WIC program was contributing to increasing health department work load in other clinic areas led to a separate analysis of encounter data for infants who were born since the data system was implemented. This is the group of those children who would not have attended a Health Department clinic prior to July 1, 1977. The data indicate not only that WIC clinic visits account for an increasing proportion of all encounters for these infants, but that more and more infants who first visit the health department for WIC certification later appear for a well-child or other clinic encounter. While it is impossible to prove that these infants would not have visited the Children's Clinic anyway, all of the encounter data strongly suggest that these infants and children who come to the Health Department for WIC certification contribute disproportionately to the increasing activity in all other aspects of Columbus County Children's Clinic operations.

TABLE 5. Non-WIC Nursing Encounters by Clients on WIC

	Total Non-WIC Encounters	Number With Clients on WIC	Per Cent
April–June 1978	213	40	19%
July–Sept. 1978	143	33	23
Oct.–Dec. 1978	355	86	24
Jan.–March 1979	333	85	26
April–June 1979	320	108	34

In order to evaluate further the findings that WIC may account for much, if not most, of the increased demand on the Children's Clinic, Columbus County information was compared with information from two North Carolina counties without WIC programs. Counties without WIC were chosen for comparison according to similarities in six criteria: geographic location, total population, infant mortality rate, percentage under five years of age, percentage nonwhite and percentage of AFDC.‖

---

§ The active case load is the total number of infants and children whose WIC coupons were collected from the Health Department in a given month. Data are from the WIC program of the Columbus County Health Department.

‖ The source of this information is the North Carolina Department of Administration, Division of Budget and Management.

The two counties chosen, Beaufort and Duplin, were found to have had increases in the total nursing units of service for infants and children of 7 and 16 per cent, respectively.# This compares with an increase in nursing encounters in the Columbus County Children's Clinic of 34 per cent during the same period. Even if all WIC certification and recertification encounters with nurses are eliminated, Children's Clinic nursing activity would still show an increase of 25 per cent.

A number of fortuitous circumstances facilitated the Health Department's response to this increased demand. Following the opening of a new county hospital, the Health Department moved into the old hospital building. The availability of new space permitted the Children's Clinic to increase its operations from two to five days per week. Not only were the two full-time child health nurses able to spend more of their time in the Clinic, but a part-time nurse was hired to help out. In recognition of the contribution of the Health Department to child health in Columbus County, the County Commissioners increased local contribution to the Maternal and Child Health budget 20 per cent between FY78 and FY79. However, no special recruitment program was undertaken to increase WIC enrollment. The principal means for informing the public remained word of mouth.[7]

Since January 1979, a change in WIC reporting requirements has made it slightly easier for health departments to request funds for nursing staff equivalents. However, this can alleviate only part of the problem. WIC funds can only be used for reimburse for WIC-related services. The increasing percentage of children on WIC receiving non-WIC services from public health nurses suggests that these same children may be contributing disproportionately to the increase in all nursing encounters. If this is the case, consideration must be given to providing the additional resources necessary to handle this load. New mechanisms for supporting traditional child health activities in health departments taking on a WIC program may be indicated.

Acknowledgments

This report would have been impossible without the cooperation of the Columbus County Health Department, and especially of John R. Black, M.D., Health Director. The Children's Clinic staff who provided all the services and data were Ruth Storms, P.N.P., Marjorie Council, R.N., Hildred Scarborough and Ruth Long. WIC case-load data were provided by Judith Long and Sandra Roholt. Finally, none of the data would have been interpretable without the technical assistance and consultation of Sharon Schramm.

References

1. Foreman C. Hearings before the Subcommittee on Elementary, Secondary, and Vocational Education. U.S. Congress, House, Committee on Education and Labor, 92nd Congress, 2nd session, April 27, 1978; 184.

2. Edozien JC, Switzer, BR, Bryan RB. Medical Evaluation of the Special Supplemental Food Program for Women, Infants, and Children. Chapel Hill, N.C.: University of North Carolina, School of Public Health, Department of Nutrition, 1976.

3. Weiler PG, Stalker HP, Jennings SW, et al. Anemia as a criterion for evaluation of a special supplemental food program for women, infants, and children. Pediatrics 1979;63:584.

4. Food and Nutrition Committee. Report to the Urban Affairs Council, 1969. Cited in To Save the Children: Nutritional Intervention through Supplemental Feeding. U.S. Congress, Senate, Select Committee on Nutrition and Human Needs, 93rd Congress, 2nd session. Washington, D.C.: U.S. Government Printing Office. 1974:3.

5. Humphrey HH. Congressional Record. U.S. Congress, Senate, 92nd Congress, 2nd session, August 16, 1972;S28588.

6. U.S. Congress, Senate, Committee on Agriculture, Nutrition, and Forestry. The Child Nutrition Amendments: Hearings before the Subcommittee on Nutrition, on S-3805. 95th Congress, 2nd session, April 6, 1978;187–240.

7. Black JR. Personal commuinication, May 4, 1981.

# Maternal and Child Health Quarterly Reports, Maternal and Child Health Branch, North Carolina Division of Health Services.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Closure Among U.S. Community Hospitals, 1976–1980

A Description and a Predictive Model

ROSS M. MULLNER, PH.D.,* CALVIN S. BYRE, PH.D.,*
PAUL S. LEVY, SC.D.,† AND JOSEPH D. KUBAL, M.A.*

Closure of U.S. community hospitals during the years 1976–1980 is analyzed by describing the distributional patterns of closings among hospitals classified by certain institutional characteristics: bed size, number of facilities and services, ownership, teaching status, location in a Standard Metropolitan Statistical Area (SMSA) or in a non-SMSA, and location in one of the nine United States Census Divisions. The method of Cox regression analysis is used to show that each of these institutional characteristics is independently, and significantly, associated with closure, and to estimate the strength of the association. On the basis of this analysis, a theoretic model is constructed for estimating the relative probability of closure for any community hospital with a given set of these institutional characteristics.

THERE IS A GROWING AWARENESS among researchers that hospital closure in the United States is closely and significantly related to events and processes occurring in the health care industry in general, and that an understanding of hospital closure is requisite for fully understanding and dealing with such larger issues as the problem of reimbursement, the capability of health care institutions to adapt to changes in the communities they serve, excess hospital bed capacity, equal access to good quality health care and the distribution of physicians. For this reason, hospital closure has in recent years come to figure prominently in an increasing number of studies.[1] Intrinsic studies of closure are comparatively rare, however, apart from several studies that examine a few particular cases; these cases are instructive as examples but furnish evidence that is difficult to generalize. Yet it is clear that a solid body of descriptive information about closure, showing where and when it most frequently occurs, and a theoretic model capable of predicting its occurrence, are necessary before the implications of closure for other health care issues can be adequately explored and before effective policies can be devised for dealing with it.

One descriptive study on a fairly large scale was that of Ruchlin, Pointer and Cannedy, who in 1972 and again in 1973 examined the characteristics of nonfederal short-term hospitals registered with the American Hospital Association (AHA) that closed in the United States in 1969.[2–3] They compared the distributions of closed hospitals with those of hospitals that remained open within categorizations of hospitals by location in U.S. Census Divisions, by types of facilities and services offered, by bed size, and by control and

---

* Hospital Data Center, American Hospital Association, Chicago, Illinois.

† School of Public Health, University of Illinois at the Medical Center, Chicago, Illinois.

Address for reprints: Ross Mullner, Hospital Data Center, American Hospital Association, 840 N. Lake Shore Drive, Chicago, IL 60611.

0025-7079/82/0700/0699/$01.05 © J. B. Lippincott Co.

ownership (i.e., for-profit, state and local government, and voluntary). They also calculated "the mean value of select operating characteristics of failed and nonfailed hospitals" in order to isolate the variables that seemed most indicative of "operational viability." However, while their methologic approach was fresh and stimulating, their findings and conclusions were somewhat questionable and of limited usefulness. Of the 55 hospital closings that occurred in the single year that they considered, they were only able to deal with the 23 for which data were available. Moreover, they did not attempt to discover the relative strength of the association of the variables they examined with closure; hence, their model for predicting institutional viability is not capable of determining the relative probability that hospitals with given values for these variables will close.

McNeil and Williams, in "Wide Range of Causes Found for Hospital Closures,"[4] dealt with 231 U.S. hospitals registered with the AHA that closed between January 1975 and December 1977. The descriptive categories they used were similar to those used by Ruchlin et al. The focus of their article, however, was the causes of hospital closure in the U.S. In investigating this, their approach was not to construct an objective model of closure based on salient features of closed hospitals, but rather to classify and interpret the reasons given by hospital personnel themselves for the closings of their hospitals.

Mullner, Byre and Kubal, in "Closed U.S. Community Hospitals, 1972–1978: Perspectives and Trends,"[5] examined the 315 U.S. community hospitals registered with the AHA that closed within a seven-year period. This study was primarily descriptive, analyzing the distributions of closed hospitals within categories similar to those used by Ruchlin et al. and by McNeil and Williams, and noting how they vary from the distributions of all hospitals within the same categories.

It is the purpose of the present study to continue the lines of inquiry pursued in the three articles just mentioned. The present study employs similar categories, based on certain institutional characteristics, to describe the distributional patterns of closings among the AHA-registered community hospitals that closed in the U.S. within the five-year period from 1976 to 1980. It focuses on them because they form a large and relatively homogeneous class of hospitals (they included 5,851, or 83.4 per cent, of the 7,015 U.S. hospitals existing in 1978, the middle year of the period). Community hospitals are defined by the AHA as "all nonfederal short-term general and other special hospitals, excluding hospital units of institutions, whose facilities and services are available to the public."[6] Furthermore, this study uses multiple regression analysis to measure the strength of the association of closure with each of the institutional characteristics considered and, on the basis of this analysis, constructs a theoretic model for estimating the relative probability of closure for any community hospital with a given set of these characteristics.

The set of institutional characteristics by which hospitals might be classified and whose association with closure might be analyzed is vast. A subset whose members are well defined, determinate, objectively quantifiable and relatively stable has been selected for examination; previous research indicates that these characteristics have an important relation to closure. They are bed size, number of facilities and services, control (investor owned; nongovernment, not-for-profit; and government, not-for-profit), teaching status, location in a Standard Metropolitan Statistical Area (SMSA) or in a non-SMSA, and location in one of the nine United States Census Divisions.

This article is divided into two main parts. The first part begins with a brief descriptive overview of closure among all U.S. hospitals during the years from 1976 to 1980, dividing them into the two classes of community hospitals and noncommunity hospitals. The latter is a catchall, heterogeneous class comprising federal hospitals (psychiatric and general and other special) and nonfederal hospitals (including psychiatric hospitals, tuberculosis and other respiratory disease hospitals, and long-term general and other special hospitals). The rest of this part is devoted to closure among community hospitals. The second part performs the multiple regression analysis and constructs the predictive model.

All data used in this study are drawn from the AHA Membership, Publication, and Circulation File.

Description

Overview of Hospital Closure in the U.S., 1976–1980

Table 1 shows the number of U.S. hospitals registered with the American Hospital Association that closed and the number of beds lost as a result, by year and by classification as community or noncommunity. In general, closures and beds lost to closures were most numerous toward the beginning of the period, after which they declined, only to increase again at the end. A total of 296 hospitals closed together with their 29,491 beds. Although most of these hospitals were community hospitals (226), the bed losses were fairly evenly divided between the two classes, indicating that closed noncommunity hospitals were on the average much larger than closed community hospitals.

Community Hospital Closures, 1976–1980

Closings by Bed-Size Category. Table 2 shows for each bed-size category of community hospitals the number of hospitals and beds lost to closings during the five years and the proportion of all closed hospitals and closed beds that they represent. Calculated for each category is the rate of closure and of beds lost thereby per 100 hospitals and 100 beds in existence in 1978, the middle year of the period. For purposes of comparison, the number of community hospitals and beds existing in each category in 1978 is also shown, with the proportion of all community hospitals and beds that each represented.

It is apparent from Table 2 that, the smaller the bed-size category, the more subject its members are to closure: the rates of both closings and of their resulting bed losses decrease progressively with increases in bed size, dropping precipitously between the 100–199 and the 200–299 bed-size categories. Closure is, in fact, overwhelmingly a phenomenon of hospitals with fewer than 199 beds; making up 71.5 per cent of the nation's community hospitals and containing only 35.3 per cent of their beds in 1978, they accounted for 95.1 per cent of all community hospitals that closed during the five years and for 80.0 per cent of their bed losses due to closure.

Closings by Control Category. Table 3 shows the number, proportions, and rates of closings and resulting bed losses in each of the three control categories of community hospitals. The number of community hospitals and beds existing in each category in 1978 is shown, as well as the proportion of all community hospitals and beds that each represented.

Investor-owned (for-profit) community hospitals constitute the smallest category, with 12.5 per cent of all community hospitals and 8.3 per cent of their beds in 1978, but they accounted for the largest number of both closings (110, or 48.7 per cent) and of beds lost (6,876, or 42.2 per cent). Nongovernment, not-for-profit community hospitals, on the other hand, which make up the largest category, with 57.1 per cent of all community hospitals and 70.1 per cent of their beds in 1978, accounted for only 34.9 per cent (79) of closings and for only 38.3 per cent (6,228) of beds lost. State and local government community hospitals, the second largest group, with 30.4 per cent of hospitals and 21.6 per cent of beds in 1978, also had fewer closures (37, or 16.4 per cent) and somewhat fewer beds lost (3,171, or 19.5 per cent) than might be expected. The rates show similar relation-

TABLE 1. Closed Hospitals and Resulting Bed Losses by Hospital Type and Year of Closure

Hospital Type	Number of Closed Hospitals and Resulting Bed Losses											
	Year of Closure											
	1976		1977		1978		1979		1980		Total	
	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds
Community	52	3926	62	3943	37	2464	33	2721	42	3221	226	16,275
Noncommunity	18	4431	13	1660	16	2766	9	1687	14	2672	70	13,216
Total	70	8357	75	5603	53	5230	42	4408	56	5893	296	29,491

TABLE 2. Closed Community Hospitals and Resulting Bed Losses by Bed-Size Group

Bed Size	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978				Rate of Hospital Closure per 100	Rate of Bed Closure per 100
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Per Cent of Total Beds		
6– 24	30	13.3	511	3.1	268	4.7	5,077	0.5	11.2	10.1
25– 49	83	36.8	2,833	17.4	1,075	18.3	39,136	4.0	7.7	7.2
50– 99	62	27.4	4,218	25.9	1,444	24.7	103,460	10.6	4.3	4.1
100–199	40	17.7	5,463	33.6	1,398	23.9	196,652	20.2	2.9	2.8
200–299	7	3.1	1,546	9.5	717	12.2	173,751	17.8	0.9	0.9
300–399	2	0.9	698	4.3	392	6.7	132,828	13.6	0.5	0.5
400–499	1	0.4	432	2.7	240	4.1	106,352	10.9	0.4	0.4
500+	1	0.4	574	3.5	317	5.4	218,147	22.4	0.3	0.3
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7

ships and point to the same conclusion: for-profit community hospitals are by far the most subject to closure of the control categories, and are followed in the far distance by state and local government and by nongovernment not-for-profit hospitals.

Closings by Hospital Location. Table 4 shows the numbers, proportions, and rates of closings and resulting bed losses among community hospitals classified first according to location in the nine U.S. Census Divisions and then by location in SMSAs or non-SMSAs. The SMSA is defined by the U.S. Office of Management and Budget as a "county or group of contiguous counties that contain one or more cities of at least 50,000 population or twin cities with a combined population of 50,000 or more."

The rates of closures and bed losses reveal how disproportionately closings were distributed among the Census Divisions. The Pacific region, with rates of 7.0 for hospitals and 3.5 for beds, was the region whose community hospitals were most subject to closure. Having 12.5 per cent of all U.S. community hospitals and 10.9 per cent of their beds in 1978, it accounted for 23.0 per cent of their closings and 23.2 per cent of their bed losses through closure during the five-year period. The mid-Atlantic region also is noteworthy for having high rates and for disproportionately large numbers of closures and resulting bed losses. It contained 10.9 per cent of all community hospitals and 17.2 per cent of all community hospital beds in 1978, but accounted for 16.7 and 28.9 per cent of all community closures and beds lost, respectively. The West South Central region is interesting precisely because it is comparatively unremarkable: it had a slightly larger number of community hospital closures (37, or 16.4 per cent of the total) and a slightly lower number of beds lost to closures (1,421, or 8.7 per cent of the total) than might have been expected, given that it had 14.6 per cent of all community hospitals and 10.3 per cent of community hospi-

TABLE 3. Closed Community Hospitals and Resulting Bed Losses by Hospital Control

Hospital Control	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978					
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Per Cent of Total Beds	Rate of Hospital Closure per 100	Rate of Bed Closure per 100
Nongovernment—not-for-profit	79	34.9	6,228	38.3	3,339	57.1	683,258	70.1	2.4	0.9
State and local government	37	16.4	3,171	19.5	1,780	30.4	211,099	21.6	2.1	1.5
Investor-owned (for-profit)	110	48.7	6,876	42.2	732	12.5	81,046	8.3	15.0	8.5
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7

TABLE 4. Closed Community Hospitals and Resulting Bed Losses by U.S. Census Region and SMSA/Non-SMSA Location

Census Region SMSA/Non-SMSA Location	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980): Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Total U.S. Community Hospitals and Beds Existing in 1978: Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Per Cent of Total Beds	Rate of Hospital Closure per 100	Rate of Bed Closure per 100
New England	7	3.1	468	2.9	259	4.4	51,015	5.2	2.7	0.9
Mid-Atlantic	37	16.4	4,736	29.1	636	10.9	167,869	17.2	5.8	2.8
South Atlantic	28	12.4	1,877	11.5	809	13.8	149,590	15.3	3.5	1.3
East North Central	23	10.2	1,809	11.1	910	15.6	193,224	19.8	2.5	0.9
East South Central	16	7.0	765	4.7	484	8.3	68,019	7.0	3.3	1.1
West North Central	24	10.6	1,369	8.4	799	13.7	98,972	10.2	3.0	1.4
West South Central	37	16.4	1,421	8.7	853	14.6	100,785	10.3	4.3	1.4
Mountain	2	0.9	62	0.4	362	6.2	39,424	4.1	0.6	0.2
Pacific	52	23.0	3,768	23.2	739	12.5	106,505	10.9	7.0	3.5
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7
SMSA	148	65.5	12,620	77.5	2,953	50.5	725,444	74.4	5.0	1.7
Non-SMSA	78	34.5	3,655	22.5	2,898	49.5	249,959	25.6	2.7	1.5
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7

tal beds in 1978. This is the region that, in the late 1960s and in the early and mid-1970s, accounted for the greatest number and the highest proportion of closed community hospitals.[3,4,5]

All other regions had fewer closings and resulting bed losses than might have been expected from the proportion of U.S. community hospitals and beds that they contained. The Mountain region had not only the fewest closings, but also the lowest proportion of hospitals and beds lost to closings.

As the rates of bed losses due to closings indicate, metropolitan and nonmetropolitan areas did not differ significantly in the proportions of their community hospital beds lost through hospital closings. There was, however, a significant difference in the proportions of their hospitals that closed: metropolitan hospitals, which made up half (50.5 per cent) of all community hospitals in 1978, accounted for 62.1 per cent of the closings.

Closings by Number of Facilities and Services. Table 5 shows the numbers, proportions, and rates of closings and resulting bed losses among community hospitals classified by numbers of facilities and services. As the columns with the rates show, by far the highest proportions of closings and of bed losses occurred among hospitals with only one to four facilities, and the proportions became lower as bed size increased.

Closings by Teaching/Nonteaching Status. Table 5 also shows the distributions of closings and of resulting bed losses among community hospitals classified according to their status as teaching or nonteaching hospitals. The AHA considers a hospital to be a teaching hospital if it meets any one of the following criteria: 1) it has a medical residency program approved by the American Medical Association; 2) it is a member of the Council of Teaching Hospitals; or 3) it is affiliated with a medical school. As can be seen from the table, teaching hospitals are much less subject to closure than are nonteaching hospitals. While the overall U.S. rate for hospital closure was 3.9, and for bed losses due to closure 1.7, the rates were only 1.1 and 0.6, respectively, for teaching hospitals, but 4.6 and 3.2 for nonteaching hospitals.

Discussion

In the preceding description of the distributions of closure in the various categorizations of community hospitals, it was assumed, as it was assumed by earlier researchers, that each of the institutional characteristics on which the categorizations are based bears a significant and unique associative relationship to closure. It was assumed, that is, that each characteristic affords a perspective onto a different and separate facet of the phenomenon.

It will be evident upon reflection, however, that this assumption is not necessarily correct. Through examination of the relationship between closure and each characteristic taken singly, the possible relationships that may obtain between the various characteristics themselves have not been taken into account. To take an example, among hospitals classified by control, the highest rates of closure were found among the investor owned; among hospitals classified by size, among the smaller bed-size groups. But investor-owned hospitals tend on the average to be smaller than those of the other control types; the high rate of closure among them may be due, therefore, not to some property that they possess by virtue of their form of control, but to some property that they possess by virtue of their size.

The validity of the assumption must, therefore, be tested. In order to discover whether, and how strongly, each of the institutional characteristics that has been considered is independently associated with closure, a multiple regression analysis was performed to obtain estimates of the relationship between closure and each

TABLE 5. Closed Community Hospitals and Resulting Bed Losses by Number of Facilities and Services and Teaching Status

Number of Facilities and Services, Teaching Status	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978					
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Percent of Total Beds	Rate of Hospital Closure per 100	Rate of Bed Closure per 100
—*	3	1.3	58	0.4	586	10.0	52,889	5.4	0.5	0.1
1– 4	69	30.5	2,033	12.4	312	5.4	10,221	1.1	22.1	19.9
5– 9	68	30.1	3,930	24.0	1,187	20.3	65,748	6.7	5.7	6.0
10–14	58	25.8	6,000	36.6	1,210	20.7	119,712	12.3	4.8	5.0
15–19	17	7.5	2,090	12.7	920	15.7	146,508	15.0	1.8	1.4
20–24	8	3.5	1,127	6.9	633	10.8	156,318	16.0	1.3	0.7
25+	3	1.3	1,155	7.0	1,003	17.1	424,007	43.5	0.3	0.3
Total	226	100.0	16,393	100.0	5,851	100.0	975,403	100.0	3.9	1.7
Teaching	14	6.2	3,464	21.1	1,278	21.8	567,660	58.2	1.1	0.6
Nonteaching	212	93.8	12,929	78.9	4,573	78.2	407,743	41.8	4.6	3.2
Total	226	100.0	16,393	100.0	5,851	100.0	975,403	100.0	3.9	1.7

* Hospitals shown on this line were nonresponders to the AHA Annual Survey of Hospitals; data on their facilities and services are not available.

characteristic, while controlling for confounding by the other characteristics. The data source used for the analysis was the 1976 AHA Annual Survey Tape; it lists all U.S. hospitals in operation as of September 30, 1975, the time immediately preceding the period covered by this study and at which all closed hospitals examined here were still open.

From this source and from the information on closed hospitals contained in the AHA Membership, Publication, and Circulation File, a working tape was constructed containing the following variables for each of the 5,859 hospitals in operation as of September 30, 1975:

Number of beds;
Whether or not it was a teaching hospital;
Whether or not it was located in an SMSA;
Number of facilities and services (0–52);
Whether or not it was investor owned;
Whether or not it was controlled by a state or local government;
Location by U.S. Census Division; and
Whether or not it closed between Jan. 1, 1976 and Dec. 31, 1980 and, if it closed, the month.

The procedure followed in the multiple regression analysis was developed by D. R. Cox and is commonly known as "Cox regression" or "life table regression."[7,8] It is now widely used by epidemiologists and biometricians for analyzing survival time as a function of a set of independent variables when the maximum time in which any individual can be studied is determined by the length of the study period. The method is appropriate in the present instance, in which the "individual" is the hospital and the study period is five years. The model specified by the procedure assumes that the chance of a hospital's closing during a particular period is a product of two functions, one depending only on time and the other depending only on the characteristics of the hospital. The model is specified as follows:

$$h(t) = \lambda(t)\, f(X_1, \ldots X_k), \qquad (1)$$

where h(t) = the chance of a hospital's closing in t years, $\lambda(t)$ = the function of time, and $f(X_1, \ldots X_k)$ = the function of the hospital characteristics X_1, X_k.

The function $\lambda(t)$ is common to all hospitals, regardless of their characteristics. It is the function $f(X_1, \ldots X_k)$ that determines the effect of hospital characteristics on the risk of closure and that is, therefore, the regression equation to be estimated. The form of this function most often used in the Cox regression is the exponential given by:

$$f(X_1, \ldots X_k) = \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k), \quad (2)$$

where $\beta_1, \ldots \beta_k$ are the regression coefficients to be estimated. They are estimated by the statistical procedure of *conditional likelihood* developed by Cox; this makes it possible to estimate the regression coefficients without explicitly formulating the function $\lambda(t)$.

The variables were specified as shown in Table 6. Number of beds and number of facilities and services are the only quantitative variables; the others are categoric. Location in U.S. Census Divisions has been coded as two dichotomous variables (the usual procedure in multiple regression analysis for categoric variables having more than two categories). The remaining variables are already dichotomous.

The variables were fitted to the exponential equation (2) by means of the Cox regression computer program known as PHGLM, which has recently been incorporated into the Statistical Analysis System (SAS) and is described in the SAS Supplemental Library Users Guide.[9]

Table 7 shows the magnitude of the estimated Cox regression coefficient for each variable and the statistical significance of the relationship between the variable and closure when there is control for possible confounding due to the other variables. As can be seen, each of the characteristics examined in the descriptive analysis does in fact have a significant independent association with closure. Thus, the Cox regression analysis validates the descriptive categories employed in this study.

TABLE 6. Variables

Variable	Coefficient	Specification
Number of beds (X_1)	β_1	Actual number of beds
Teaching status (X_2)	β_2	1 if teaching hospital; 0 if nonteaching hospital
Number of facilities and services (X_3)	β_3	Actual number of facilities and services
Investor owned (X_4)	β_4	1 if investor owned; 0 otherwise
Government controlled (X_5)	β_5	1 if controlled by a state or local government; 0 otherwise
Mid-Atlantic or Pacific location (X_6)	β_6	1 if located in Mid-Atlantic or Pacific Census Division; 0 otherwise
Mountain location (X_7)	β_7	1 if located in Mountain Census Division; 0 otherwise
SMSA location (X_8)	β_8	1 if located in an SMSA; 0 otherwise

The Theoretic Model

Not only does the Cox regression analysis validate the use of these institutional characteristics to form descriptive categories for investigating hospital closure, but, as will be seen, the technique can be expanded to create a theoretic model for estimating hospitals' relative risks of closing. First, odds ratios were calculated that estimate the risk of closure for hospitals having a given level of each variable (e.g., location in an SMSA) relative to hospitals having some reference level of the same variable (e.g., location in a non-SMSA) when the values of the other variables are the same. They were derived from the model as follows:

$$R = e^{\beta(X_1 - X_2)} . \qquad (3)$$

TABLE 7. Estimated Regression Coefficients and Relative Risks as Derived from Cox Regression Model

Variable	Specification	β Coefficient	Significance (p value)	Relative Risk Odds-Ratio Derived from Model
Beds	100 beds/300 beds	−0.0075	<0.0001	4.51
Teaching status	Teaching/nonteaching	1.0332	0.0008	2.81
Investor owned	Yes/no	1.0278	<0.0001	2.79*
SMSA	Yes/no	0.7993	0.0000	2.22
Coastal location	Mid-Atlantic and Pacific/other region	0.6389	0.0000	1.89†
Facilities	4 facilities/10 facilities	−0.0769	<0.0001	1.58
State or local government control	Yes/no	−0.4828	0.0231	0.62*
Mountain location	Mountain/other region	−1.5787	0.0272	0.21†

* Relative to nongovernment not-for-profit.
† Relative to nonmountain and noncoastal.

where R = relative risk of closure, X_1 = specified level of a variable X, X_0 = some reference level of a variable X, and β = regression coefficient for X as obtained from Cox regression. The relative risks estimated in this way control for confounding by the other variables being investigated.

The relative risk odds ratios shown in Table 7 indicate that, for example, a hospital having 100 beds has a 4.5 times greater risk of closure than does a hospital having 300 beds, *ceteris paribus*, an investor-owned hospital, 2.79 times greater than a nongovernment not-for-profit hospital, etc. Relative risks lower than 1.00 indicate a protective effect. For example, a hospital that is not government controlled has a 1.6 times greater risk of closure than does a government-controlled hospital (i.e., 1/0.62 = 1.61). Here it may be noted that, while the significance of the association noted above between teaching status and closure was expected, the *nature* of that association is somewhat surprising. It was shown in the descriptive analysis—in which there was no control for confounding effects—that teaching hospitals are less likely to close than are nonteaching hospitals. It is seen, however, that teaching status per se is associated with a higher likelihood of closure than is nonteaching status. This finding of an unexpected positive association should be regarded with caution, however, since six of the 14 teaching hospitals that closed were located in the New York City SMSA. That the association between teaching status and closure is positive rather than negative may be a fact that is peculiar to the present data set, due to unusual circumstances affecting the New York hospital industry during the period; it might not be the case in other years.

Next, the regression coefficients estimated as described above were used to calculate a score, S, for each of the 5,859 community hospitals considered in the study, based on the particular values of its characteristics, by means of the formula:

$$S = \beta_1 X_1 + \ldots, + \beta_k X_k. \tag{4}$$

The correlations between the scores, S, of the hospitals and their rates of closure were then examined by grouping the hospitals into deciles on the basis of their scores and computing the five-year closure rates for each decile. As Table 8 shows, the scores generally increase with increases in the closure rates; this correlation is especially evident in the deciles with the higher scores and is most striking in the

TABLE 8. Prognostic Scores of Closed Hospitals 1976–1980 Compared With Number of Closings and Closure Rates

Decile*	Prognostic Scores(S)	Number of Closures	Five-year Closure Rate (per 100 Hospitals)
1st	−17.05 to −3.69	1	0.17
2nd	−3.68 to −2.73	3	0.51
3rd	−2.72 to −2.19	3	0.51
4th	−2.18 to −1.75	4	0.68
5th	−1.74 to −1.42	9	1.54
6th	−1.41 to −1.15	5	0.85
7th	−1.14 to −0.84	19	3.24
8th	−0.83 to −0.48	28	4.78
9th	−0.47 to 0.22	38	6.47
10th	0.23 to 2.37	116	19.83

* Five hundred eighty-six hospitals in all except the 10th decile, which contains 585 hospitals.

tenth decile, in which a five-year closure rate of 19.83 per cent is found among hospitals with scores of 0.23 to 2.37.[‡]

Thus, the score obtained for any community hospital can be interpreted as a prognostic index of its relative likelihood of closure.

Conclusion

This study has described the distributions of closed community hospitals within categories of hospitals classified according to the institutional characteristics of bed size, control, Census Division location, SMSA/non-SMSA location, number of facilities and services, and teaching status. It has demonstrated the validity of the assumption that these characteristics are intrinsically related to closure by means of the technique of Cox regression analysis. It has constructed from this method of analysis a theoretic model for estimating the relative probability of closure for any hospital having a given set of the characteristics.

It was not the purpose of this study to determine what the necessary or sufficient conditions of closure are, nor to discover its causes. It can be assumed that the closure of a hospital is typically the outcome of a long, dynamic process involving a number of complex, interrelated factors in the internal system of the institution (e.g., an inadequate number of medical staff, financial distress, underutilization, etc.) and in its external operating environment (e.g., competition from other institutions, changes in the population of the community served, etc.). The isolation of these and the discovery of the concatenations of causes and effects that are most often operative in hospital closure remain an important task for health care researchers to accomplish. Prerequisite for its accomplishment are, first, a comprehensive body of general information about the scope and dimensions of the phenomenon and, second, a series of prospective case studies examining the phenomenon *in process* in a large number of hospitals over a long period of time. It is hoped that this descriptive analysis has added to the former; the authors believe that this theoretic model, which makes it possible to identify those hospitals that are most likely ultimately to fall victim to closure, will facilitate the latter.

Acknowledgments

The authors would like to express their gratitude to Dale Matthews, Project Director, Hospital Data Center, American Hospital Association, and to Norman Iverson, Programmer Analyst, University of Illinois School of Public Health at the Medical Center, Chicago, for their expert assistance in writing the computer programs needed for this study.

References

1. Institute for Health Planning. Capacity reduction through facility closure and conversion: a bibliography with abstracts. Prepared for the Health Resources Administration, HRA 232-79-0036. Madison, Wis.: Institute for Health Planning, April 1980;6-18.

2. Ruchlin HS, Pointer DC, Cannedy LL. Hospitals that fail are often the small ones. Mod Hosp 1972;120:61.

3. Cannedy LL, Pointer DD, Ruchlin HS. Viability and hospital failure: methodological considerations and empirical evidence. Health Serv Res 1973;8:27.

4. McNeil D, Williams R. Wide range of causes found for hospital closure. Hospitals 1978;52:76.

5. Mullner R, Byre C, Kubal J. Closed U.S. community hospitals, 1972–1978: perspectives and trends. Soc Sci Med 1980;14D:355.

6. American Hospital Association. Hospital Statistics, 1980 Edition. Chicago: American Hospital Association, 1981;ix.

7. Cox DR. Regression models and life tables. J R Statist Soc Series B 1972;34:187.

8. Lee ET. Statistical methods for survival data analysis. Belmont, Cal.: Lifetime Learning Publications, 1980;306-18.

9. Harrell F. The PHGLM procedure. In: SAS Supplemental Library Users' Guide 1980 Edition. Cary, N.C.: SAS Institute, Inc, 1980;119-31.

[‡] Teaching status was included in the calculation of the scores shown, although, as mentioned above, the regression coefficient that was calculated for that variable might be valid only for the present data set. The scores were also calculated while leaving teaching status out of account, but they did not change appreciably.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Self-Care Education

Impact on HMO Costs

DONALD W. KEMPER, MSIE, M.P.H.*

The ability of a medical self-care program to have an impact on the utilization and cost of in-clinic and referral visits in a prepaid group practice was studied by following the medical encounter records of 900 HMO members in 217 families during a 12-month period. A control group selected by randomization after acceptance of an initial invitation to participate was employed. In spite of the program's impressive effect on both self-care knowledge and per-visit costs, the study could not demonstrate any significant program effect either on the frequency or on the total costs of clinic visits.

HEALTH EDUCATION is a mandated service for federally qualified Health Maintenance Organizations (HMOs). In setting that requirement, Congress contended that to maintain good health and optimal use of health care services, a person must be motivated and knowledgeable in making health-related decisions. Although most HMOs employ enough educational activity to meet minimum regulatory levels, few have embraced the spirit with which the requirement was imposed.

HMO managers have been reluctant to make substantial investments in educational strategies without a base of factual evidence indicating bottom-line cost reductions.

The impact of a medical self-care education program on the frequency and cost of health care services received by its participants in a prepaid group practice HMO is examined. Medical self-care education in this article refers to general educational efforts addressing the person's need to make decisions and apply home treatments regarding a broad range of common minor illnesses and injuries. The evaluation employs a control group selected by randomization after acceptance of an initial invitation to participate.

Background and Previous Findings

Although the overall self-care/self-help movement has many varied and early roots in the United States,[1,2] Keith Sehnert's *Course For the Activated Patient*[3] was the first of a new generation of medical self-care programs to address a broad range of health concerns and a basic change in the patient's role in the doctor/patient relationship. In the ten years since its 1970 introduction, similar classes and workshops have been organized in virtually all states of the nation, and self-care books, guides and videotapes have been developed to help meet the public's increasing interest in regaining self-responsibility for their health.[3–13]

Early evaluation of the Activated Patient program with both Medicaid and Medicare participants was hampered by methodologic and procedural problems

---

* Healthwise, Incorporated, Boise, Idaho.

Supported jointly by Grant R18 HS 02109 from the National Center for Health Services Research and by a grant from the W. K. Kellogg Foundation.

Reprint requests: Donald W. Kemper, Executive Director, Healthwise, Incorporated, PO Box 1989, Boise, ID 83701.

0025-7079/82/0700/0710/$00.95 © J. B. Lippincott Co.

that neither confirmed nor refuted its value.[4] The evaluation did suggest, however, that people would enroll and attend the classes and that their confidence about self-care increased.

Similarly, separate evaluations of 1975 self-care classes in rural Idaho and Oregon communities revealed substantial improvements in self-care knowledge and high enrollment and attendance levels.[14–15] Conclusions were positive and optimistic, but no direct measures of program effects on cost or utilization were obtained.

More recently, an evaluation of a self-care program for university students with upper respiratory infection symptoms suggested substantial savings in both visits and costs due to the program.[16] Other studies of categoric self-care programs involving asthma patients and home throat culture kits have contributed further to the evidence that self-care efforts can yield credible returns.

Two studies that evaluate the impact of the popular self-care book *Take Care of Yourself*[12] on physician use rates have been reported. Two physicians at the University of Washington compared patients' actual decisions to see a doctor with the actions recommended by algorithms in the book.[17] For the eight symptom categories studied, strict adherence to the algorithms would have increased the number of visits over that observed for five categories. Decreased utilization was predicted only for cases of hoarseness. As noted by the researchers themselves, generalizations based on the results are limited.

A second evaluation of *Take Care of Yourself* involved a prospective randomized control trial in a prepaid health plan.[18] The study followed the physician utilization of 1,300 families divided into three groups. Group 1 was the control group. Group 2 members received the book and were invited to a one-hour self-care seminar. Group 3 received the book, the seminar invitation and a $50 refund if the family's office visits dropped by at least one third during the six-month trial period The study found widespread use of the book within groups 2 and 3, and a larger decrease in clinic use rates for those groups than for the controls. However, statistical significance of the impact could not be demonstrated.

Although these previous attempts to evaluate self-care education in HMOs and other settings have generally returned positive and optimistic conclusions, they have fallen short of the direct financial evidence needed to justify large scale investment in general (as opposed to categoric) medical self-care initiatives. The present study was designed to help fill that need.

Description of the Program

The self-care program involved a series of ten two-hour workshop sessions led by a nurse practitioner. The format was informal, with emphasis placed on skill building and peer reinforcement through demonstrations and discussions. Active participation in the workshops was encouraged as the key method of developing the confidence needed to apply the skills and knowledge presented. Greater confidence in dealing with health concerns was viewed as equal to greater self-care competence in the overall purpose and goals of the program.

The program content (see Fig. 1) began with vital sign- and symptom-monitoring skills. Participants were taught to perform home physical examinations on themselves and their children. Subsequent sessions dealt with the management of abdominal problems, backaches and headaches, upper respiratory problems, skin problems and minor injuries. The topic selection was based on the frequency of visits reported in the National Ambulatory Medical Care Survey[20] and on participant interest, expressed in previous versions of the program. Discussions also emphasized

FIG. 1. Ten topics on family health.

- The Home Physical Exam
- Abdominal Health Problems
- Backaches and Headaches
- The Common Cold
- Upper Respiratory Infections
- Skin Problems
- Injuries
- Emergencies
- Dental Care & Nutrition
- Life Style and Health

a more activated relationship between the person and his/her health care provider.

Later in the series, the topics shifted away from minor illness and injury. Participants were told that although practical skills in those areas may save them money and improve their self-confidence in dealing with health problems, long-term improvements in health would depend more on broader life-style changes. Accordingly, ~~the~~ last few workshop sessions focused more on stress management, nutrition and fitness.

Each participant was given a 250-page self-care guide, *The Healthwise Handbook*,[8] and assigned readings prior to each session. The guide became a continuing reference source after the workshops were over. Consistency among workshop groups was maintained through the use of an instructor's guide[5] and a series of videotapes supporting the content and concepts of the program. Distribution of the support materials has allowed program replication in more than 15 states.

Evaluation Methods

Members of a small (6,000-person enrollment) prepaid group practice health plan in Boise, Idaho were contacted first by letter and then by phone to learn of their interest in attending a new health education class to be offered. Those who said yes (48 per cent) were randomly divided into experimental and control groups. Experimental group members were then encouraged to attend the actual workshop. Controls were asked to complete a questionnaire and told that although the trial workshops were filled, they would be recontacted if and when additional classes were offered. As shown in Figure 2, 59 individuals (55 per cent of the experimental group) actually participated in at least one workshop. Because of the program's emphasis on childhood and young adult problems, only those families with children under 18 years of age were asked to participate.

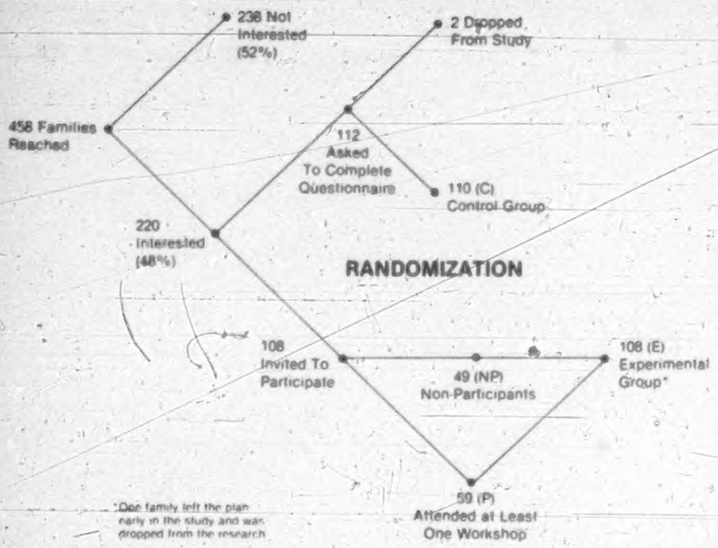
Although only 59 of the 107 families actively participated, the entire experimental group was used in the analysis to assure equivalency with the controls (Fig. 2). Demographic comparisons between the two groups revealed no apparent anomalies (Table 1).

The evaluation covered three general concerns: 1) the logistic success of the program; 2) the program impact on the predisposing factors of self-care knowledge and attitudes; and 3) the program impact on clinic utilization and cost.

Logistic success was measured by enrollment penetration rates, attendance and participant ratings of the workshop.

Self-care knowledge gains were measured by a 69-item self-care test designed around the specific learning objectives of each workshop session. The instrument was refined through several trial tests and professional reviews before being used. A split-half reliability test using the Spearman–Brown prophecy formula resulted in an R value of 0.94.[21]

Data on the utilization and cost of services by both the experimentals and controls were obtained from a computerized summary of clinic and referral encounters during a 12-month period. Cost data were based on the nonmember fees of the HMOs for the services received. The data


458 Families Reached
238 Not Interested (52%)
2 Dropped From Study
112 Asked To Complete Questionnaire
110 (C) Control Group
220 Interested (48%)
RANDOMIZATION
108 Invited To Participate
49 (NP) Non-Participants
108 (E) Experimental Group*
59 (P) Attended at Least One Workshop

*One family left the plan early in the study and was dropped from the research

FIG. 2. Experimental and control group selection.

were part of the routine financial management system of the HMO. Encounters were assigned to the three categories of preventive, minor or other, based on the symptom/problem codes. The International Classification of Health Problems in Primary Care (ICHPPC) system was used.[22] The "minor" category was designed to include most of the problem classifications that the self-care program directly addressed.

Analysis of the utilization and cost data was based on the statistic

$$X_i = \frac{\text{number of visits by family of participant } i \text{ during the 12-month period}}{\text{number of family members in family } i}$$

The t-test was used to test for significant differences.

Findings

Findings in the first two areas of evaluation were positive and highly consistent with results from previous studies discussed earlier. After adjusting for the control group, 26 per cent of families contacted had one or more members who participated in the workshop. Those who did participate attended an average of 5.5 of the ten sessions.

Eighty-three per cent of participants said the workshop improved health care for their families, and 81 per cent reported they had read at least half of the handbook.

Self-care knowledge scores (Table 2) increased 125 per cent for the participants and dropped only 1 per cent during the following six months. Scores for the control group increased only 8 per cent during the same period. The score improvement difference between the participants and controls was highly significant at the 0.01 level ($t = 9.53$).

Impact on Utilization

Findings regarding the program impact on utilization frequency of clinic services were less definitive. As indicated in Table 3, the experimental group averaged slightly more visits to the HMO clinic than did the controls. However, when referral visits are considered, the total frequency of visits is slightly higher for the controls. These differences are small and do not approach statistical significance. (The largest t value was for referral visits, $t = 1.26$.)

Additional analysis of the utilization data by calendar quarter, by demographic subgroups and for children only demonstrated neither statistical significance nor strong trends. A Kolmogrov–Smirnov test[23] did not refute the hypothesis that the two populations are identical.

Impact on Cost

The analysis of financial impact of the program included study of both costs per family member and cost per visit.

As indicated in Table 4, both the participating and nonparticipating elements of the experimental group accrued less cost over the study period than did the controls. Although the average cost differences reached as high as 24 per cent for referral and "other" clinic services, the high variances encountered prevented any statistically convincing findings. Only for the 42 per cent difference in referral cost averages between workshop participants and controls was statistical significance approached (t = 1.654).

Findings in the cost-per-visit analysis were more definitive. As shown in Table 5, the $2.54 per-visit difference between experimentals and controls was found to be significant at the 0.01 level (t = 2.67). When the per-visit costs were broken down into clinical, laboratory and x-ray components, only the $1.19 difference related to laboratory costs remained significant (t = 2.55). The apparent 28 per cent savings in laboratory cost and the 24 per cent savings in x-ray accounted for 73 per cent of the overall cost/visit margin.

As in the cost-per-visit-person analysis, it was found that both the experimental participants and the non-participants showed significantly lower costs/visits than did the controls (t = 2.06 and 2.16, respectively).

The overall cost per family of in-clinic and referral costs as shown in Table 6 reveals an average cost savings of $55.48 per family, or $5,975 for the experimental group. However, since only 59 of the 107 experimental group families participated in the workshop, the total average savings per participant was $101.27. It should be noted that the cost data applies to the 12-month period from December 1977 through November 1978.

TABLE 1. Comparison of Experimental and Control Groups, Prepaid Group Practice

	Experimental	Control
Number completing demographic questionnaire*	70	93
Age (mean)	35.4	36.1
Education (mean years of schooling completed)	14.5	14.8
Children (number under 18 years of age)	2.3	2.3
Sex		
Per cent female	86	72
Per cent male	14	28
Income		
Per cent < 5,000	0	1
5,000–9,999	8	11
10,000–14,999	37	33
15,000–19,999	19	26
Per cent > 19,999	29	24
Unknown	13	5
Pretest		
Knowledge (mean score)	28.3	27.3
Attitude (mean score)	7.7	7.0
Satisfaction (mean score)	7.0	6.8

* Questionnaire response rate averaged 75 per cent. Because utilization and cost data were available on all of the randomized subjects, the analysis in those areas was based on a larger sample.

Program Cost

Excluding program development and research expenditures, the program cost totaled approximately $65 per participating family. As shown in Table 7, 58 per cent of the cost was due to personnel and indirect costs, while direct expenditures for workshop supplies and handbooks amounted to $17 per family. Although some economies of scale would result from large-scale implementation of the program, the total cost would not be significantly changed.

TABLE 2. Self-Care Knowledge Scores

Self-Care Knowledge Scores	Experimental Group		Control Group
	Participants	Non-participants	
Pretest (n)	28.5 (59)	27.1 (11)	27.3 (93)
Posttest (n)	65.1 (40)	NA	NA
Per cent change*	124% ↑	NA	NA
Six-month test (n)	66.0 (34)	31.2 (9)	30.2 (71)
Per cent change from pretest†	111% ↑	8% ↑	8% ↑
Per cent change from posttest†	1% ↓	NA	NA

* Comparison of pretest to posttest scores of only those returning both questionnaires.

† Comparison of pretest to six-month and posttest to six-month scores of those returning all questionnaires.

Discussion

The primary question regarding self-care programs is whether, in the long run, their participants benefit, either by better health or by lower medical expenses. This study was unable to address the first half of the question directly because of resource limitations and the difficulty of assessing health-status changes. However, the cost issue was addressed, at least in part, and a number of secondary and logistic questions were answered with relevance to both cost and health-status concerns.

Logistically, it was demonstrated that large numbers of people from a general clinic population could be recruited and trained in medical self-care at a reasonable cost. This is a necessary but certainly not sufficient condition for program cost effectiveness. However, by enrolling more than 25 per cent of the HMO families contacted at a single point in time, the project reaffirmed the opportunity that self-care programs have for affecting large segments of the population.

Although health care knowledge and awareness are not directly related to health status, they are generally accepted as positive predisposing factors that health improvement programs should address. The study results clearly reinforce earlier findings that program participants do learn substantial amounts through self-care workshops and classes. In addition, these findings indicate that the retention of such knowledge can be high, at least for the six months following the program's completion.

TABLE 3. Visits/Family Member/Family/Year*

Visit Category	Experimental E			Control (C)	t (C–E)
	Participants	Nonparticipants	Combined (E)		
Number of families	59	48	107	110	
In-clinic visits					
Minor	2.68	2.05	2.40	2.22	−0.72
Preventive	0.87	0.63	0.76	0.75	−0.06
Total	4.31	3.67	4.03	3.99	−0.08
Referral visits	0.28	0.37	0.322	0.423	1.26
Total visits	4.59	4.04	4.35	4.41	

* $\dfrac{\text{Number of visits by family of participant } i \text{ during 12 months}}{\text{Number of family members in family } i}$

TABLE 4. Cost/Family Member/Family/Year

Cost Area	Experimentals			Controls (C)	Significance Tests							
					E vs C		P vs C		NP vs C		P vs NP	
	Participants (P)	Nonparticipants (NP)	Combined (E)		Per Cent Change	t	Per Cent Change	t	Per Cent Change	t	Per Cent Change	t
In-clinic visits												
Minor	45.50	36.74	41.57	38.48	+8.0	—	+18.2	—	-4.5	—	+23.8	—
Preventive	19.21	12.53	16.21	17.19	-5.7	—	+11.7	—	-27.1	—	+53.3	—
Other	23.65	21.85	22.84	30.31	-24.6	—	-22.0	—	-27.9	—	+8.2	—
Total	88.35	71.11	80.62	85.98	-6.2	—	+2.8	—	-17.3	—	+24.2	—
Referrals	28.49	47.88	37.19	49.23	-24.5	1.015	-42.1	1.645*	-2.7	—	-40.5	1.13
Totals	116.85	118.99	117.81	135.21	-12.9	—	-13.6	—	-12.0	—	-1.8	—
Number of families	59	48	107	110								

* Significant at 0.1 level.
t values less than 1.00 omitted.

TABLE 5. Cost/Visit/Family by Cost Element

Cost Element	Experimentals			Controls (C)	t = Test Values			
	Participants (P)	Nonparticipants (NP)	Combined (E)		E vs C	P vs C	NP vs C	P vs NP
Clinical cost/visit	16.07	15.13	15.64	16.34	-1.17	-.313	-1.71	+1.05
Lab cost/visit	3.01	3.22	3.10	4.29	-2.55*	-2.20*	-1.67	-.41
X-ray cost/visit	1.87	2.39	2.10	2.77	-1.35	-1.48	-.60	-.76
Total cost/visit	20.96	20.74	20.86	23.40	-2.67*	-2.06*	-2.16*	+.18
N	59	48	107	110				

*Significant at 0.01 level.

The utilization data were both unexpected and enlightening. Based on the earlier findings by Moore and the program emphasis placed on reducing inappropriate utilization, a substantial reduction of in-clinic visits had been expected. It had been predicted that little impact would be seen in the specialty referral areas. In fact, the reverse appears more true. The observed reduction in referrals for the participants may be explained by one or more of the following speculations:

1. Better self-care applied earlier in the course of an illness led to fewer complexities of the illnesses encountered; thus, fewer specialty services were needed.
2. The program further developed the participants' confidence in their primary care professionals so that demands for referrals were reduced.
3. Statistical aberration—a larger sample would have refuted the findings (the finding was not statistically significant).

Although the favorable findings in the reduction of per capita ambulatory costs are not statistically convincing, the program impact on per-visit costs is. The definitive finding that self-care classes lead to lower laboratory costs per visit is, however, difficult to explain. Although workshop content includes encouragement of patient questioning of the purposes of laboratory tests and x-rays, large cost differences in those areas were not expected.

Also difficult to explain are the consistent differences between the experimental nonparticipants and the controls. Since neither group received any more "treatment" than a brief program description, the differences must be explained by a basic population bias. The classic randomization of the initial experimental and control groups would suggest that experimental group individuals declining to participate tended to use fewer and less expensive services than those who chose to come to the workshops. Such a self-selection bias would tend to strengthen the cost-effectiveness benefits of the program.

TABLE 6. Cost per Family

Cost Area	Experimental ($)	Control ($)	Savings ($)
In-clinic visits	333.05	353.33	20.28
Referrals	147.39	182.95	35.56
Total	480.44	536.28	55.84
Number of families	107	110	

Total savings = 107 × 55.84 = $5,974.88; savings/workshop participant: $5,974.88 ÷ 59 = $101.27.

Conclusions

It can be concluded from the findings that the self-care program caused a reduction in the cost per visit incurred by program participants. Further, it was shown that more that 25 per cent of the young families in a clinic population voluntarily participated in the program and that their knowledge of the appropriate management of common health problems increased significantly.

Although both clinic and referral costs were lower for the experimental group than for the controls, the high variance encountered prevents conclusions either de-

TABLE 7. Program Replication Cost

Instructors	
Three nurse practitioners (20 class hours each)	$ 875
59 Healthwise Handbooks, $11* each	649
59 sets of supplies, $6 each	354
3 playings of Healthwise videotapes, $1,000 × 30%	300
10 days of health educator's time to coordinate classes, $70	700
Recruitment costs (estimated)	300
Total direct costs	$ 3,178
Indirect support cost (20%)	636
Total cost	$ 3,814
Total cost per participant	$ 64.64

* Now available at quantity discount for $4.64.

nying or confirming a favorable impact on total costs.

Recommendations

The findings of this study point directly to the urgent need to develop improved self-care education programs and to evaluate more fully their effect on health status and economic benefits:

1. Future research into the impact of self-care programs on clinic utilization should attempt to reduce the effect of high variance on the analysis by consideration of family size and previous health history in the design of the study statistic. (Example: $x = b$ visits $+ b_1$ family size $+ b_2$ health status.)
2. Sample size should be adequate to demonstrate statistical significance if the sample means differ by 10 per cent or more.
3. Long-term evaluations covering three or more years should be attempted, to measure any delayed changes in knowledge, attitude and utilization.
4. Stable clinical settings should be selected to minimize both the number of out-of-clinic visits and the turnover of the medical staff.

References

1. Green LW, Werlin SH, Schauffler HH, Avery CH. Research and demonstration issues in self-care: measuring the decline of mediocentrism. Health Educ Mongr Summer 1977.

2. Levin LS, Katz AH, Holst E. Self-Care Lay Initiatives in Health. New York: Prodist, 1976.

3. Sehnert K, Nocerino JT. The Activated Patient: A Consumer-Oriented Program on Preventive Medicine and Self-Help Medicine. Washington, D.C.: The Center for Continuing Health Education, Georgetown University, 1977.

4. The Center for Continuing Health Education. A Course and Evaluation of The Activated Patient Concept. Washington, D.C.: The Center for Continuing Health Education, Georgetown University, 1977.

5. Healthwise, Incorporated. Healthwise Instructor's Guide. Boise: Healthwise, Incorporated, 1979.

6. Georgetown University. Medicaid Citizens as Activated Patients: A Course Guide. Washington, D.C.: Georgetown University, September 1976.

7. Reeves J. Health Skills for Parents. Health Skills Associates, 1975.

8. Roberts T, Tinker K, Kemper D. Healthwise Handbook. Boise, Idaho: Healthwise, Incorporated, 1979.

9. Samuels M, Bennett H. The Well Body Book. Clinton, Mass.: Colonial Press, Inc, 1973.

10. Sehnert K. How To Be Your Own Doctor (Sometimes). New York: Grosset & Dunlap, 1975.

11. Sehnert K. The Patient as a Paramedical. Virginia Medical Monthly 1972;99:409.

12. Vickery DM. Take Care of Yourself: A Consumer's Guide to Medical Care. Reading, Mass.: Addison–Wesley Publishing Company, 1976.

13. Vickery DM, Fries JF, Pantell RH. Taking Care of Your Child. Reading, Mass.: Addison–Wesley Publishing Company, 1977.

14. Grant RH. Family and self help education in isolated rural communities. Health Educ Monogr 1977;5:145.

15. Kemper, DW, Tinker K. Common Sense–Common Health Evaluation. Boise: Healthwise, Incorporated, 1975.

16. Apka J, Averill BW. Self care for colds: a cost-effective alternative to upper respiratory infection management. Am J Public Health 1979;69:814.

17. Berg AO, LoGerfo JP. Potential effect of self-care algorithms on the number of physician visits. N Engl J Med 1977;300:535.

18. Moore SH. Randomized Controlled Trial of a Self Care Book. MPH thesis, University of Washington, Seattle, Washington, 1978.

19. Green LW, Drenter MW, Deeds SG, Partridge KB. Health Education Planning. Palo Alto, Calif.: Mayfield Publishing, 1980;8.

20. National Center For Health Statistics. National Ambulatory Medical Care Survey, Preliminary Data From the 1973–1974 Survey, July 15, 1975.

21. Downie NM, Health RW. Basic Statistical Methods. New York: Harper & Row, 1974.

22. World Organization of National Colleges, Academics and Academic Associations of General Practitioners and Family Physicians. International Classification of Health Problems in Primary Care. Chicago: American Hospital Association, 1975.

23. Blalock HM. Social Statistics. New York: McGraw–Hill, 1960.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Reimbursing for Residency Training

How Many Times?

MOHAN L. GARG, SC.D.,* MOUNIR ELKHATIB, M.D.,† WARREN M. KLEINBERG, M.D., M.P.H.,‡ AND JACK L. MULLIGAN, M.D.§

Through an investigation of charges generated for 105 patients treated for congestive heart failure, three issues of reimbursement policy for graduate medical education are examined: 1) reimbursement policy by third parties for resident salaries and for the portion of teaching physicians' time devoted to graduate medical education; 2) the level of involvement for physicians in dual roles as attending physicians for private patients when simultaneously engaged in graduate medical education and supervising patient care delivered by residents; and 3) the greater utilization of hospital resources in teaching units. The results demonstrate that the reimbursement policy, as presently practiced, ignores the basic differences in the level of involvement between teaching and nonteaching physicians. The data also show for the first time the incremental effect of graduate medical education in a teaching hospital. Several alternatives are suggested to eliminate the deficiencies.

IN RESPONSE TO FISCAL PRESSURES, policymakers are seeking areas in which to contain health care costs. The Health Care Finance Administration, the federal agency responsible for Medicare reimbursement, has started to look at reimbursement policies for physicians involved in teaching and patient care. However, precise data to determine the differences in level of involvement between teaching and private physicians have not been reported in the literature. Some data suggesting excessive utilization of ancillary services and increase in length of stay for patients under the care of residents have been reported.[1,2] The two studies have neither a homogeneous patient population for comparison nor a control on attending physicians.

A study was carried out to compare the components of cost of congestive heart disease patients treated by residents under supervision of teaching physicians in a teaching unit and those treated by attending physicians in a private unit of the same hospital. The study looked at two issues: 1) the greater number of diagnostic tests ordered by the medical resident; and 2) the level of involvement of teaching physicians in comparison with that of private physicians for managing the same type of patient. Both issues have implications for the effect of a graduate medical education

---

* Department of Medicine, Medical College of Ohio, Toledo, Ohio.

† Department of Family Medicine, Medical College of Ohio, Toledo, Ohio.

‡ Department of Pediatrics, Medical College of Ohio, Toledo, Ohio.

§ University of Missouri School of Medicine at Kansas City.

Supported by a grant from the National Fund for Medical Education, Hartford, Connecticut.

Reprint requests: Mohan L. Garg, Sc.D., Division of General Internal Medicine, Department of Medicine, Medical College of Ohio, C.S. # 10008, Toledo, OH 43699.

0025-7079/82/0700/0719/$00.90 © J. B. Lippincott Co.

program on hospital rate setting in teaching hospitals.

Under the present reimbursement policy for hospital care by third-party insurers (Medicare, Medicaid and Blue Cross), the salaries and fringe benefits of residents and fellows, and the time devoted by the teaching physician to supervise the graduate medical education are considered general operating expenses of the hospital. Stated another way, physician employee expenses tend to increase per unit cost of each of the services produced by the hospital. For example, the daily room rates for the intensive care unit, coronary care unit, and semiprivate and private rooms include a specific amount that is generated because of teaching programs. In addition, the relative charges of producing some ancillary services also include a teaching-generated component. In short, every unit of service cost produced in a teaching hospital contains a portion that is due to the existence of teaching programs. This has been considered to be equitable by the Graduate Medical Education National Advisory Committee (GMENAC)[3] and the Institute of Medicine (IOM)[4] because residents purportedly spend 80 per cent or more of their time in patient care activities.

Materials and Methods

The data for the present study were collected from a community hospital associated with a medical school. All patients with a diagnosis of congestive heart disease during the 12 months of 1978 were selected for the study. Patient bills were gathered from the business office, and all diagnostic tests and procedures, along with their charges, were recorded. Charges for pharmacy and other supply items were also recorded. The data were then correlated with the medical charts to determine the day tests were ordered and the source of each order, whether a resident, attending physician or standing order of the hospital. Patients were also identified by location, the teaching unit or the private unit. All of the data were compiled for each patient ($n = 105$), each patient being given a coded identification number to assure confidentiality. The data were then keypunched and stored in the computer for subsequent analysis.

Unlike previous studies, strict controls on patient comparability were maintained, and all patients had a primary diagnosis of congestive heart failure at discharge confirmed by the medical records. Other patient characteristics such as sex, age and mode of payment were gathered to examine the group for homogeneity. Severity was controlled for by the method of the New York Heart Association Criteria Committee.[5]

Results

Of 105 patients with congestive heart disease there were 52 men and 53 women. Eighty-three were 65 years of age and over, and 22 were under the age of 65. Forty patients were treated in a teaching unit, 65 in a private unit. All of the patients in the private unit had associated secondary conditions, and 95 per cent of those in the teaching unit had secondary conditions. There was no significant difference in severity of congestive heart failure or of the associated conditions between study groups, nor was there a difference in mode of payment between groups.

Table 1 presents a summary comparison of teaching versus private unit patients for various categories of hospital services use. On an average, charges for diagnostic tests are 52 per cent higher for teaching patients, $1,414 for patients in the teaching unit as compared with $932 for patients in the private unit. The differences vary widely among laboratory departments, microbiology being almost three times higher and respiratory therapy and cardioperfusion being two times higher for teaching unit patients. In Table 2, average

TABLE 1. Average Charges for Patients With Congestive Heart Disease in Teaching and Private Units

	All Patients	Type of Patient: Private	Type of Patient: Teaching	Per Cent Increase or Decrease in Charges
Number of patients	105	65	40	
Average charge ($)				
Total	2969	2820	3210	+14
Room	1641	1704	1538	−10
Pharmacy	134	113	169	+50
Supplies	78	72	89	+24
Diagnostic tests	1116	932	1414†	+52
Hematology	50 (5.6)*	46 (5.3)	56 (6.1)	+22
Microbiology	40 (3.5)	23 (1.9)	68 (6.0)	+196
General chemistry	322 (20.5)	266 (15.9)	411 (28.0)	+55
Cardiovascular	129 (3.8)	136 (3.7)	118 (4.1)	−13
Radiology	123 (5.4)	108 (4.8)	148 (6.4)	+37
Respiratory therapy and cardio-perfusion	304 (10.5)	218 (8.0)	444 (14.4)	+104
Miscellaneous	148 (6.3)	135 (5.4)	166 (7.7)	+23

* Average number of tests.

† Figures may not total accurately due to rounding error.

charges per patient for teaching and private unit patients with the same attending physicians are shown. Even though the attending physicians are the same, diagnostic and treatment charges are 64 per cent higher in the teaching unit. The three most excessively utilized areas in a teaching unit are microbiology, cardiovascular, and respiratory therapy and cardio-perfusion.

Since all of the patients have comparably severe congestive heart failure, the explanations for the differences in treatment must be found in the source of the treatment orders. The data in Table 3 show the

TABLE 2. Average Charges per Patient With Congestive Heart Disease for Private and Teaching Units With Same Attending Physician

	Private Unit	Teaching Unit	Per Cent Increase (+) or Decrease (−) in Charges
Number of Patients	20	21	
Average total charges ($)	2912	3716	+28
Room	1673	1714	+2
Pharmacy	131	219	+67
Supplies	94	112	+19
Diagnostics and Treatment	1014 (51.7)*	1661 (83.2)	+64
Hematology	49 (5.9)	65 (7.0)	+33
Microbiology	41 (3.6)	92 (8.2)	+124
General chemistry	277 (17.7)	406 (28.6)	+47
Cardiovascular	57 (3.1)	141 (4.1)	+147
Radiology	174 (5.1)	162 (7.5)	−7
Respiratory therapy and cardio-perfusion	239 (9.3)	621 (19.6)	+160
Miscellaneous	177 (7.0)	174 (8.2)	−2

TABLE 3. Distribution of Percentage of Tests Ordered by Orderer in Teaching and Private Units

Service Unit	Physician Group	Number of Patients	Per Cent of Tests Ordered by: Total	Attendee	Resident	Other
All patients	All	105	100	48	40	12
Teaching	All	40	100	23	69	9
	1	21	100	26	69	9
	2	19	100	18	71	12
Private	All	65	100	73	12	15
	1	20	100	75	13	12
	3	45	100	72	12	16

distribution of tests by order for patients in each unit. As expected, in the teaching unit residents order 69 per cent of all tests and attending physicians order 23 per cent. This contrasts with the private unit, where residents order only 12 per cent of all tests, while attending physicians order 73 per cent. The residents have supervised primary responsibility for the teaching unit and are allowed to write orders with later consultation on teaching rounds. On the private unit, they can write orders only with the attending physician's direct prior approval. In Table 3, group 1 physicians are those who have patients in both teaching and private units (same as in Table 2). Yet the data imply that although they are the attending physicians in both units, the physicians are much more involved in the care of their patients on the private unit as compared with the teaching unit.

To examine the pattern of involvement on a day-to-day basis, Figure 1 illustrates the proportional ordering pattern of diagnostic tests for the first nine days of hospital stay among teaching and private unit patients. The data clearly indicate that attending physicians' involvement is virtually complete among their own private unit patients, as is expected. Among teaching unit patients, attending physicians are involved much less frequently, as demonstrated by the proportion of diagnostic tests ordered, on no day being responsible for a majority of care as measured by the proportion of tests ordered. The residents are the

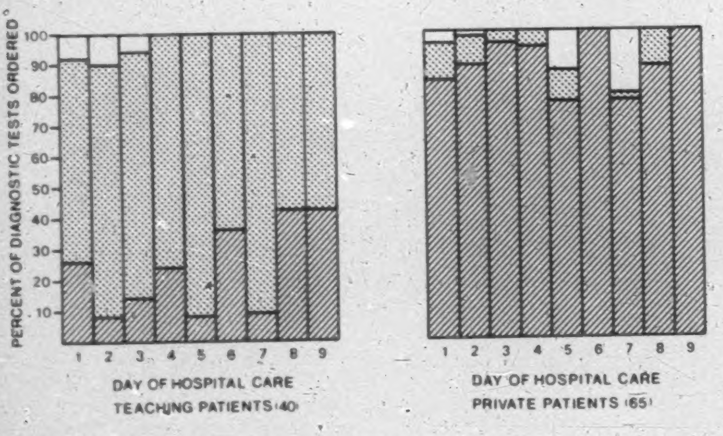


FIG. 1. Per cent distribution of diagnostic tests ordered by attending physician and resident; and standing orders for teaching and private service patients. (Striped area, attending physician; stippled area, resident; blank area, standing.)

primary communicators to patients and their families on the teaching service, whereas private attending physicians do most of the talking to patients and their families on the private service.

Figure 2 presents a comparison of ordering patterns for attending physicians (group 1 physicians from Table 3) who had patients in teaching as well as in private units. Here, also the same pattern persists, and the relationship of the attending physician in the private unit as much stronger. For private unit patients, attending physicians initiate virtually all orders, and residents play little role. Attending physicians in the teaching unit, however, are on no hospital day more involved with their patients than are the residents.

Discussion

The results of the present study raise several issues concerning reimbursement policy for graduate medical education: 1) reimbursement policy by third parties for resident salaries and for the portion of teaching physicians' time devoted to graduate medical education; 2) the level of involvement for physicians in dual roles when they are attending physicians for private patients, and simultaneously engaged in graduate medical education and supervising the patient care delivered by residents; and 3) the amount of extra utilization of hospital resources in teaching units.

Under the present method of reimbursement by third parties, the resident salaries and fringe benefits, and the part of teaching physicians' salaries for supervising graduate medical education are permitted to be included in the general operating expenses of a teaching hospital. The arrangement leads to a somewhat uniform increase in charges for each unit of service produced for patient care and subsequently billed to third parties. Thus, the charge for each unit of service includes a portion that relates directly to the teaching program. Under this arrangement every patient is paying for the cost of medical education, at a uniform rate. There is a charge differential between a hospital that is involved with a graduate medical education program and another that is not.

In a teaching unit, the care is primarily delivered by the resident under the supervision of a teaching physician. It is also apparent from two earlier studies[1,2] that the utilization of hospital services is greater in teaching units. In the present study, in spite of a control on primary diagnosis and on supervisory physicians, the utilization is 64 per cent higher in teaching units than in private units.

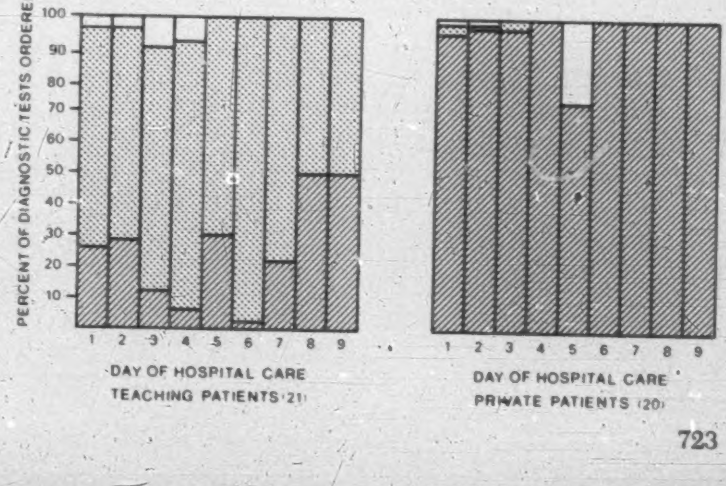


FIG. 2. Per cent distribution of diagnostic tests ordered by attending physician and resident, and standing orders for teaching and private service patients with same group of attending physicians. (Striped area, attending physician; stippled area, resident; blank area, standing.)

In two earlier studies[6,7] on the purpose of the tests ordered by residents,[6,7] it was found that approximately 10 per cent of the tests were ordered for defensive and educational purposes. Of course, residents are likely to order extra tests because they are learning medicine, and, as their experience increases, they are likely to be more careful in ordering.

The increase in number of diagnostic tests brings with it another increase in interpretation fees for excessive testing. For example, in the private unit, an average of 2.5 EKGs per patient were ordered, as compared with 3.6 in a teaching unit. Thus, the interpretation fee for 1.1 EKGs will be extra in a teaching unit.

The progressive increases in the cost of patient care for the three types of patients are shown in Figure 3. The figure illustrates the number of items that should be considered when estimating the total cost of graduate medical education and that were not meant to be quantified in the study. The first bar in Figure 3 illustrates that, in a nonteaching hospital, there are generally three components of patient charges for inpatient care: 1) hospital charges; 2) interpretation fees of support physicians, such as pathologists, radiologists and other consultants; and 3) fees of attending physicians. The second bar represents the private unit of a teaching hospital; there are added charges due to the existence of teaching programs. These charges include only the salaries and fringe benefits of the residents, and the portion of salary and fringe benefits of the teaching physicians. In the last bar, which represents the teaching unit of the teaching hospital, there are two additional items that increase the charges as compared with those of a private unit. The items are overutilization of diagnostic tests and procedures, and additional reimbursement for services of support physician services, such as pathology, radiology and cardiology, as a consequence of overutilization.

The study uncovers two key dilemmas in medical education. The first dilemma is documented by the excessive utilization and the teaching physician's salary added to the charges in the teaching unit. The second derives from the basic process by which resident education in hospitals is financed.

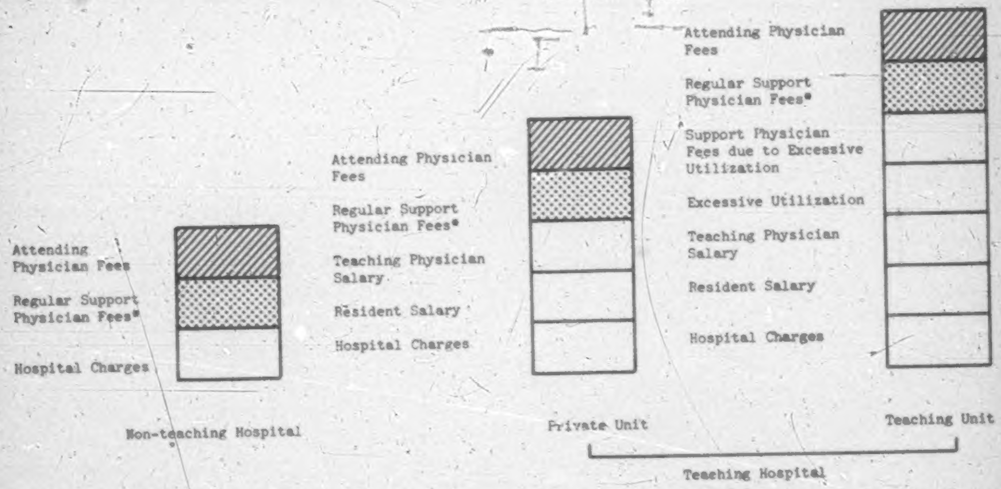


FIG. 3. Progress increase in charges for patients with congestive heart disease as a consequence of graduate medical education.

* Interpretation fees of pathologists, radiologists, cardiologists, etc.

Two methods of education are used in current residency training programs: one method utilizes learning by doing under guidance and supervision, whereas the other represents learning from a role model. A difference is seen in the levels of control exercised by teaching physicians over patient care on the private unit compared with the teaching unit. When teaching physicians exercise close control of patient care, the cost of the health care is clearly lower. Two levels of health care cost, if not of health care quality, exist in many teaching hospitals, and the difference raises questions of ethics and public policy. For example, in 1976 there were 15,381 full-time physicians in hospitals belonging to the Council of Teaching Hospitals[8] and 31,342 in other associated hospitals out of a total of 343,876 professionally active physicians. As the present analysis indicates, the level of involvement in patient care is different, but the level of reimbursement is the same. The policy is based on the premise that teaching physicians should be reimbursed equally if "personal and identifiable"[9] services were provided. However, the present data indicate the contrary. Therefore, the reimbursement policy, as presently practiced, ignores the basic difference in the level of involvement. Obviously, if the reimbursement is appropriate for one group, then it is inappropriate for the other.

Since the cost of teaching physicians' salaries is built into hospital overhead, and subsequently into daily bed rates, the problem of teaching physicians' salaries added to their fees as attending physicians bears reevaluation. There have been no studies evaluating the loss in fees for teaching physicians due to their teaching time. Teaching physicians may actually gain more in that they have "free help" to assist them in managing patients. Private physicians are reimbursed at the same rate but must manage patients with more of their personal time.

Three alternatives stand out as solutions to the problems posed. The first reflects a maintenance of the status quo, with a recognition that some costs of medical education are logical components of health care in general and should be borne by patients and their insurance companies. The slight percentage increase cost is small compared with that of other hospital charges, and the residency program allows attending physicians to be more efficient in other aspects of their practice.

The second solution represents a recognition of the need for closer supervision to prevent excessive utilization on the teaching unit and to bring health care costs to the level of the private unit. The amount of teaching physicians' salaries would depend on the recognition of the importance of teaching time and the need to attract teachers to the teaching hospital, as well as on the increased quality of care resulting from a teaching environment.

The third solution recognizes each individual component of health care costs and supports only those essential to patient care and physicians' services to be covered by insurance or private pay. Teaching functions would be supported by community, governmental or university educational funds, and would reflect the public recognition of the need for medical education and a willingness to support it. Whereas this solution would make health care and medical education more sensitive to public need, it would also make health care and medical education sensitive to public whim and political pressure.

In many teaching hospitals, attending physicians do not receive salaries from the hospital. A dilemma arises when full-time, hospital-based physicians on salary have fees collected for them by the hospital. No dilemma exists if their salaries are totally derived by fees collected and are not reflected in hospital charges. However, there is often an incentive for physicians to obtain bonuses beyond their salaries by increasing their patient contacts and, therefore, their patient billing. This can easily be done by increasing their patient contacts through the use of residents and then billing for the resident contacts.

The study implies that as more primary responsibility is given to residents on a teaching unit, greater costs result by virtue of the residents' inexperience and uncertainty. It can be surmised that close supervision and experience improve physicians' efficiency and decrease excessive utilization. Increased supervisory responsibility of teaching physicians could produce the close supervision and exposure to experience, but requires more of the teaching physicians' time and, therefore, more dependance on salary from the teaching hospital. Whether the increased efficiency and cost savings produced by having more full-time teaching faculty in hospitals balance the cost of salaries needs to be studied.

The ultimate solution, it is hoped, depends on decisions made by informed interaction between teaching hospitals, organized medicine, medical schools, and consumers and their representatives. The reevaluation of traditional approaches, advantages and disadvantages of different systems, and the basic goals of residency programs need to be considered before drastic changes are made. Some basic research and comparative studies must be carried out to provide the information needed for informed decisions.

Finally, from the standpoint of hospital income, the factors described in the study compound and increase one another, as long as the system remains openended. As limits are placed on hospital rate increases, more pressure is placed on the hospital to eliminate teaching physicians' salaries and to encourage fee collection by current separate mechanisms. Indeed, a clear separation of physicians' fees and the hospital bill would allow for more clear delineation of cost factors. Moreover, if residents assist physicians in their work, then their salaries should be derived as a component of the physicians' fees, and not the overall hospital charges. Medical care and liability consequences would be clearly the ultimate responsibility of attending physicians. It is hoped that attending physicians would have incentive for greater supervision and improved role modeling. The possibility that attending physicians would become dictatorial would be balanced by the competition among hospitals for a limited number of residents and the need to provide a good teaching environment to attract residents.

References

1. Schroeder SA, O'Leary DS. Difference in laboratory use and length of stay between university and community hospitals. J Med Educ 1977;52:418.

2. Martz EW, Ptakowski R. Educational costs to hospitalized patients. J Med Educ 1978;53:383.

3. Department of Health, Education and Welfare. Interim Report of the Graduate Medical Education National Advisory Committee to the Secretary. Washington, D.C.: Department of Health, Education and Welfare, April 1979.

4. Medicare–Medicaid Reimbursement Policies. Institute of Medicine of the National Academy of Sciences, March 1979.

5. New York Heart Association Criteria Committee. Nomenclature and Criteria for the Diagnosis of Diseases of the Heart and Great Vessels. New York: Little, Brown, 1973.

6. Garg ML, Gliebe WA, Elkhatib M. The extent of defensive medicine: some empirical evidence. J Legal Med 1978;6(2):25.

7. Garg ML, Gliebe WA, Elkhatib M. Medical practice for educational purposes: the internal medicine resident in a hospital 1978;52(4):97.

8. COTH Director. Educational Programs and Services. Council of Teaching Hospitals, Association of American Medical Colleges, 1978.

9. Hadley J. Medical Education Financing: Issues and Options. Department of Health, Education and Welfare. [Publication No. (PHS) Washington, D.C.: 79-3261.]

MEDICAL CARE
July 1982, Vol. XX, No. 7

Effectiveness of an Ambulatory Care Telephone Service in Reducing Drop-In Visits and Improving Satisfaction With Care

CLARENCE F. STIREWALT,* MARGARET W. LINN, PH.D.,† GUSTAVO GODOY,‡ FELIPE KNOPKA, M.D.,§, AND BERNARD S. LINN, M.D.ǁ

The purpose of this research was to determine whether a specialized telephone service could cut down on unscheduled visits to ambulatory care and improve satisfaction with care. Patients who were to receive care (N = 561) were assigned randomly to a specialized telephone service or a control group. They were measured with regard to satisfaction with care before being assigned and 6 months later. The number of scheduled and unscheduled visits also were monitored for a 1-year follow-up period. The telephone service, which provided a triage system for referrals as well as a source of central contact for scheduling or for complaints, helped to improve satisfaction and reduce the number of unscheduled visits within 6 months. Such a service also may have application in other problem areas in the delivery of quality ambulatory care.

TWO EXTREMES of patient behavior have created problems in the delivery of ambulatory care. At one extreme, there are the "no-show" patients who leave holes in appointment scheduling so that the most cost-effective use of the care providers' time is diminished. At the other extreme, there are the "drop-in" patients who disrupt scheduling in the opposite manner, overloading the system so that the quality of services provided may be jeopardized. Unfortunately, these two types of patients seldom balance each other out, so that "drop-ins" appear conveniently when "no-shows" occur. Thus, the problem is compounded.

Over a number of years, in an extremely busy outpatient department, the number of patients who came without appointments had escalated to such an extent that a full-time physician had to be assigned each day just to manage these patients. A preliminary study comparing the drop-in population with clinic users who usually sought care by appointment indicated less satisfaction with care among drop-in patients. Furthermore, the drop-in patients tended

---

From the Veterans Administration Medical Center and the University of Miami School of Medicine, Miami, Florida.

* Chief, Medical Administration Service, VA Medical Center.

† Director, Social Science Research, VA Medical Center, and Professor of Psychiatry, University of Miami School of Medicine.

‡ Administrative Officer, Research Service, VA Medical Center.

§ Associate Chief of Staff for Ambulatory Care, VA Medical Center, and Assistant Professor of Medicine, University of Miami School of Medicine.

ǁ Associate Chief of Staff for Education, VA Medical Center, and Professor of Surgery, University of Miami School of Medicine.

Supported by VA Health Services Grant 310-2.

Address for reprints: Margaret W. Linn, Ph.D. (151), VA Medical Center, 1201 N.W. 16th Street, Miami, FL 33125.

0025-7090/82/0700/0739/$01.00 © J. B. Lippincott Co.

to have a history of high clinic utilization with repeated drop-in behavior patterns.

This paper describes a prospective randomized trial of a specialized telephone system designed to cut down on drop-in visits and improve patient satisfaction with care.

Several studies have determined that symptoms and ability to pay are the best predictors of the use of health services. Anderson[1] found these two factors more important than cultural and social factors. Bice and White[2] showed that perceived symptoms were the primary variable explaining utilization, not only in this country but in several other countries as well. Avnet[3] found no difference in utilization due to income when there was no financial barrier, such as when care was provided through comprehensive insurance programs. Greenlick[4] made the same observations in a study of regular subscribers to the Kaiser Foundation Plan. Likewise, Bice et al.,[5] Mechanic,[6] and Montiero[7] concluded that sociocultural differences in utilization were diminished greatly when capacity to pay for services was equalized.

The present study concerns ambulatory care provided under the auspices of the Veterans Administration, where the ability to pay was not seen as an important variable in explaining utilization, either by drop-in or scheduled appointments. The person's perception of health, however, was seen as a major factor in determining utilization. In fact, the rationale for selecting a specialized telephone system for testing was based on the assumption that it was just this perception by the patient of the need to seek immediate medical care that resulted in an inability to wait for a scheduled appointment.

Weinerman[8] believed that patient satisfaction with care influenced the degree to which medical care was sought and the extent to which advice was accepted and heeded. Stoeckle et al.[9] suggested using satisfaction data to plan and organize facilities, and to set priorities for medical education and public health programs. Several investigators[10–12] have concluded that patient satisfaction helps in understanding why people do or do not seek care. Patient acceptance often looms as a potential barrier to implementating new modes of delivering services. Donabedian[13] suggested that patient satisfaction, along with health status, was the ultimate outcome in evaluating quality of medical care. Again, the idea of a telephone service which would provide immediate access to a health system for patients who seemed to need and seek fast access to care, appeared to have the potential to influence the degree of satisfaction, which might in turn lead to modified behavior in seeking appointments for care.

The telephone's role and potential in health care has not been explored very thoroughly. In 1971, 13 per cent of patient contacts with physicians were provided by telephone.[14] Delegation of telephone responsibility to nonphysicians has been acceptable to patients.[15] Brown[16] and Ott[17] found deficiencies in history taking and accuracy of advice among house officers and health associates who were not specially trained in telephone care. However, Strain[18] found few errors and a high level of parental acceptance with a nurse who trained for this role. In special programs for chronic illness and suicide prevention, nonphysicians have been trained as primary care providers.[19–21] The telephone has been used in poison control centers as an aid in diagnosis and in offering medical consultation to rural areas.[22] Telephone service to obtain results from throat cultures and triage by telephone in emergency room care also have been described.[23,24]

The purpose of this study was to evaluate the effectiveness of a specialized telephone service in ambulatory care using a prospective controlled study for patients who dropped in for care without an appointment in regard to improving patient satisfaction with care and in reducing the number of drop-in visits.

Method

The Setting

The Miami Veterans Administration Medical Center has one of the largest outpatient loads among VA Medical Centers. Over 200,000 outpatient visits are scheduled to its 38 clinics per year. Computerized scheduling and preordering of medical records had improved the efficiency of services for patients with scheduled appointments. Patients who dropped in for their care, however, disrupted the system. Furthermore, patients had to wait while their records were being located. They also lost continuity of care through being seen by a different physician on each visit. Each day, one physician was assigned to see drop-in patients only. It was known from the pilot study that these patients often wanted answers to questions or help with problems that did not require the services of a physician. Pharmacy services, social work, dietetics, nursing, medical administration, and other services frequently were a more appropriate source of contact than the physician. Few problems actually appeared emergent enough to require immediate on-site medical attention. It seemed likely that with a telephone service patients might be influenced to schedule an appointment if a clinic visit were needed instead of just dropping in.

The Telephone Service

A clerk trained thoroughly in clinic scheduling and who was familiar with services in the hospital and outpatient department was selected to provide the telephone contacts. Protocols for telephone triage and an extensive referral network were developed for handling patients' problems. First, the clerk determined why the patient was calling. If, for example, the patient presented family problems, a referral was made to social work service. If the patient was out of medication and was eligible for a refill, the pharmacy was contacted. The same triage was followed in regard to general medical, nursing, prosthetic or medical record services. Questions dealing with eligibility, scheduling of appointments, complaints or simply the need to talk with someone were handled directly by the telephone clerk. One of the functions of the clerk was to explain services available on a scheduled basis and to offer one source where patients could call with their questions or needs. When referral was made to other services, a form was sent to that service to follow-up on their action so that the final disposition was known. If a patient was assigned to the experimental group, and therefore eligible to use the telephone service, he was introduced to the telephone operator and given a card that included the patient's name, unique number and the telephone service number. Calls were transcribed and coded for length, caller, reason for call, referral and outcome. The telephone service operated each day during clinic hours from 8:00 to 4:30.

The Research

A multidisciplinary team of investigators planned the study and met monthly to monitor its progress. Data on characteristics of patients entered into the study and types of referrals were reviewed. Randomly selected tapes of calls also were reviewed. Intake and follow-up statistics with reasons for patient loss were monitored as well. Statistical analyses were not done until the end of the study so that bias would not be entered from knowing results of unblinded comparisons.

A research assistant randomly selected patients from among those who came in for their care without appointments over a 2-year period of intake (1979-80). During these 2 years, approximately 7500 drop-in visits occurred among 3160 different patients. The method used for selection into the study was to screen drop-in patients during 2 hours of each day. The screening hours were rotated so that different hours were covered each day of a week, with the next week rotating to different hours for the days, so that within a 4-week cycle, all hours of each day had been covered. Male patients who met the criterion of living within the area covered by the clinic were eligible for inclusion in the study if they agreed to sign informed consent. Eight per cent of those selected did not agree to participate. Patients who signed informed consent were interviewed concerning demographic characteristics such as age, education, race, social class (measured with the Hollingshead two-factor index), household composition, marital status, presenting symptoms, pattern of use of medical services, distance from the clinic and self-perception of health (measured on a 5-point scale).

Patients then completed two satisfaction-with-care scales. A 45-item scale[25] that measures three areas of satisfaction was used. These areas are professional competence of physicians, personal qualities of physicians and cost/convenience of care. Statements such as "You may have to wait a little, but you can always get a doctor," were rated on 5-point strongly-agree to strongly-disagree scales. In addition, ten items that measure a more general satisfaction with care were used. The latter scale represents the super-factors of the VA Satisfaction with Care Scale, which is scored on 4-point scales.

After baseline testing, patients assigned randomly to an experimental group (i.e., eligible to use the telephone) or a control group (i.e., handled in the usual way) by their opening a sealed envelope which contained the group assignment (made with a table of random numbers). The experimental patients were given the telephone card and encouraged to call, rather than drop in, for care. Both experimental and control patients were advised that they would be contacted again in 6 months. Data were collected from their medical records to describe their diagnoses, number of drop-in visits, scheduled visits, hospitalizations during the prior 6 months and number of medications.

Six months after baseline testing, groups were retested on the same two satisfaction-with-care scales and the same information collected from their medical records describing use of scheduled and unscheduled visits and hospitalizations. An additional record follow-up was done at 12 months. Thus, the study was completed in July 1981. Data also were collected on perception of health, employment hours and overall evaluation of care. Patients in the experimental group continued to use the telephone service for as long as the study was in progress, and scheduled and unscheduled visits were recorded at 6 and 12 months for experimental and control patients.

Data Analysis

The primary outcomes of the study were satisfaction with care and number of drop-in visits. The hypothesis was that the telephone service would be effective in increasing satisfaction with care and reducing the number of drop-in visits. Data were analyzed by comparison of the outcomes of the experimental and control groups. For satisfaction-with-care ratings, the baseline scores were held constant comparing satisfaction at 6 months in a multivariate analysis of covariance as recommended by Bancroft[26] for pre–post randomized designs. Number of drop-in visits over the follow-up of 6 and 12 months was analyzed by standard analysis-of-variance techniques.

Results

The Sample

A total of 561 patients were studied, with 279 randomly assigned to the experimental group and 282 to the control group. A total of 82 per cent were successfully followed for their reevaluations concerning satisfaction with care, and all but 9 per cent were followed for determination of scheduled and unscheduled visits. The reasons for loss of data were death, moving out of the area and inability to locate patients. Comparison of baseline characteristics between those lost to follow-up and those on whom all data were complete showed no significant differences. Thus, the sample seemed to be representative of all who entered the study.

Table 1 shows the background characteristics of experimental and control patients. There were no significant differences on any of the variables describing personal characteristics. Patients averaged about 54 years of age; they had high school educations; they were lower middle class. Most were married, were residents of the area and lived about 22 miles from the clinic. They assessed their health as fair (3.2 average on a 5-point scale).

Furthermore, the experimental and control groups did not differ significantly on characteristics that described their use of services over the 6 months before they entered the study. Table 2 shows that each group averaged 2.92 and 2.46 drop-in visits. Slightly over half of their total visits were drop-in visits. Groups averaged about four diagnoses, the major ones being hypertension, diabetes, arteriosclerotic heart disease, emphysema, peptic ulcer, arthritis and skin disorders. Each group received about three medications. The control group had slightly more (although not statistically significant) hospitalizations than the experimental group.

Use of the Telephone Service

An average of 3.3 calls per year per patient in the experimental group was recorded, 86 per cent of the group using the telephone service at some time. Calls were received throughout the day on a rather even basis, with no particular hour showing higher use than any other. The duration of each call averaged 7.2 minutes (standard deviation of 7.6). Almost half of the calls (49 per cent) were handled directly by the telephone operator, the remainder (51 per cent) being referred to other sources. The most frequent referrals were for questions about the person's medical condition, which usually were referred to a nurse or physician (33 per cent). Some patients asked for appointments or when there seemed to be a need, appointments were

TABLE 1. Comparison of Backgrounds Between Experimental and Control Patients

Variables	Groups		F-Ratio
	Experimental	Control	
Age	54.05	53.13	0.47
Education (years)	12.23	12.08	0.23
Race (white)*	0.84	0.83	0.02
Social class	3.75	3.84	1.18
Married*	0.85	0.86	0.01
Total in home	1.36	1.39	0.07
Miles from clinic	21.43	22.20	0.17
Florida resident*	0.97	0.97	0.11
Self-assessed health	3.26	3.24	0.12

* Items can be read as per cent of group by disregarding the decimal point. Social class was measured 1–5 with 5 being the lowest social class. Self-assessed health was rated 1–5 with 5 being the poorest.

TABLE 2. Comparison of Experimental and Control Patients on Health Variables at Baseline

	Groups		
Variables	Experimental	Control	F-Ratio
Average number over past 6 months			
Drop-in visits	2.92	2.46	1.06
Scheduled visits	1.79	2.18	1.33
Hospitalizations	0.17	0.25	2.02
Current number			
Medications	3.26	3.09	0.78
Diagnoses	4.67	4.35	1.28

scheduled (26 per cent of the calls). About 12 per cent of the calls were referred to the pharmacy, and less than 10 per cent for each of the following: general information, cancelling appointments, complaints, questions about eligibility for care, admission to the hospital, obtaining medical records, travel to the clinic, prosthetics, transferring records and help with completion of medical forms.

Satisfaction with Care

Table 3 shows the adjusted and unadjusted mean scores at 6 months on the two satisfaction-with-care scales. There were no differences between experimental and control groups at baseline testing. At 6 months, with prescores covaried, experimental patients had more favorable scores on all three factors of the 45–item scale, with professional competency, personal qualities, and cost/convenience of care differing between the two groups at the $p < 0.05$ level for each factor, the overall multivariate level being statistically significant ($p < 0.05$) for the three factors together.

On the 10 items that represented the super-factors of the VA Satisfaction with

TABLE 3. Comparison of Experimental and Control Patients on Satisfaction with Care

	Unadjusted Postscores		Adjusted Postscores		F-Ratios	
Variables	Experimental	Control	Experimental	Control	Unadjusted	Adjusted
Three-factor scores for satisfaction with:						
Professional competence	36.03	36.99	36.02	7.13	1.63	3.26*
Personal qualities	35.15	36.71	35.15	36.56	2.04	2.89*
Cost/convenience of care	42.85	43.74	41.87	43.99	2.99*	3.78*
Multivariate F (3 scores)						1.45*
Ten-item scale for satisfaction with:						
Overall care	3.29	3.07	3.32	2.90	4.10*	4.99*
Promptness of care	2.92	2.74	3.33	2.51	3.70*	3.98*
Consideration	3.53	3.31	3.64	3.25	3.00*	3.40*
Multivariate F (10 items)						1.79*

* $p < 0.05$.
Higher scores on three-factor scores are less favorable responses.
Higher scores on the 10 items are more favorable responses.

Care Scale, three were significantly different at 6 months at the 0.05 level. Again, the multivariate level for all 10 items was significant in discriminating the two groups at the 0.05 level. The significant item discriminators, in favor of the experimental group, were overall satisfaction with care, promptness and consideration. Although the remaining items did not reach significant levels for differences between the groups, they were all in the direction of the experimental group being more satisfied with care at 6 months.

Scheduled and Unscheduled Visits

Table 4 provides data on the number of scheduled and unscheduled visits made during the follow-up of 6 and 12 months. At 6 months, the number of drop-in visits was significantly lower in the experimental group (1.61 average) as compared with the control group average of 2.11. This difference was at a 5 per cent level of significance. The number of scheduled visits did not differ significantly between experimental and control patients. However, at 12 months, the number of drop-in visits did not differ between the two groups, nor did the number of scheduled visits. The total number of visits averaged about 3.8 for both groups. However, the number of hospitalizations was significantly higher in the control group ($p < 0.05$) than in the experimental group and could have accounted for their being in the community less time during follow-up and therefore not as likely to drop in as much. Diagnoses and medications were not different between groups throughout the study.

Discussion

The hypothesis that the telephone service would improve satisfaction with care was supported. After 6 months exposure to a specialized telephone service, experimental patients had significantly more positive attitudes as measured by the three factors on the Satisfaction with Physicians and Primary Care Scale as well as on the more general 10-item Satisfaction with Care Scale. More satisfied ratings on the cost/convenience of care are easily under-

TABLE 4. Comparison of Experimental and Control Patients at Six and 12 Months Follow-up on Health Variables

Variables	Groups		F-Ratio
	Experimental	Control	
Six Months Follow-up			
Average number over past 6 months			
Drop-in visits	1.61	2.11	3.63*
Scheduled visits	2.64	2.54	1.26
Hospitalizations	0.19	0.33	2.60
Medications	3.02	3.11	0.03
Diagnoses	4.82	4.23	1.54
12 Months Follow-up			
Average number over past 6 months			
Drop-in visits	1.65	1.35	1.58
Scheduled visits	2.18	2.24	0.66
Hospitalizations	0.16	0.41	3.12*
Medications	3.12	3.26	0.14
Diagnoses	4.38	4.46	1.03

* $p < 0.05$.

stood, because the telephone service was designed to make care more convenient and access more easily obtained. Increased satisfaction in regard to the professional competence and personal qualities of the physician is less easily understood, since physicians did not operate the telephone service, nor was any intervention in the doctor–patient relationship attempted. One answer could be that the scale, although purported to measure three discrete dimensions, measured a more generalized satisfaction with care, in reality. This could be supported by the moderately high intercorrelations between the three factors, which ranged from $r = 0.46$ to $r = 0.55$.

On the other hand, more drop-in visits were correlated with more negative attitudes on all three factors at about the same moderate levels of correlation. Therefore, when the experimental group had fewer drop-in visits as a result of the telephone service, then satisfaction could be expected to improve in all of these areas. The reason that more negative attitudes toward professional competence and personal qualities of physicians were associated with drop-in visits may be from lack of continuity of care when patients come without appointments. It may be that physicians assigned to cover drop-in visits were less satisfied when they did not see their own patients and this dissatisfaction may have been transmitted to the patients, thus influencing the patient's attitude. For whatever reason, all three factors scores concerning satisfaction with care were improved significantly in experimental patients, compared with controls. Improvement is further confirmed by the use of another scale, where the overall multivariate difference favored experimental patients more than controls. Items that contributed most to discrimination between the groups (overall care, promptness and consideration) seemed highly relevant to the services provided by specialized telephone service.

The hypothesis that a telephone service could reduce the number of drop-in visits was proven after 6 months but not at 12 months. Drop-in visits went from a baseline rate of 2.9 in the experimental group to 1.6 at 6 months. The drop-in rate stayed about the same in control patients during the same time (2.4 to 2.1). Number of scheduled visits did not differ significantly between the experimental and control patients at baseline or 6 months; however, the trend was toward an increase in the number of scheduled visits for experimental patients at 6 months. This would seem to be a logical finding since one of the functions of the telephone service was to schedule visits for patients when it was determined that a visit was necessary. Thus, the total number of visits was similar for both groups with the difference being in the way they utilized services, by appointment or not. One could argue, that utilization of care was not affected by the telephone service, even though drop-in visits were reduced at 6 months. One answer is that quality of care, as measured by patient satisfaction, was improved when drop-in visits were reduced. Certainly, it is a more efficient delivery of health services when appointments are scheduled in advance. It also may be less costly. Cost will be addressed in another article.

Patients were followed in regard to their utilization patterns through a 6-month period of time. By 12 months, drop-in visits were not significantly different between experimental and control subjects. The number of drop-in visits remained about 1.6 for experimental patients and had dropped to 1.4 in controls. What appears to have occurred is that both the rate of drop-in visits and hospitalization had stabilized in experimental patients over time. What happened with control subjects is that hospitalization rates increased over time, even though control subjects did not differ significantly from experimental patients in diagnoses (kind or number) and number of medications. If diagnoses and medications can be used as indicators of severity of condition, then control subjects who were similar to experimentals were using the hospital more by 12 months. If they used the hospital more, then they had less time in the community to use out-patient services. More importantly, it has been suggested that if outpatient visits were used more effectively, then rates of hospitalization might decrease. Unfortunately, this seldom has been found in studies.[27] Perhaps patients who utilize scheduled appointsment with the same physician have less of a need for hospitalization. Furthermore, patients who are more satisfied with their care may not need additional care through hospital admission as frequently.

Overall, a specialized telephone service for ambulatory care patients was effective in improving satisfaction and reducing drop-in visits. It also may be effective over a longer follow-up in reducing hospitalization as well. Patient response to the telephone service was favorable, with about three fourths of the patients who were eligible calling an average of three times. From another standpoint, even though the telephone was effective in reducing drop-in visits, it did not eradicate the drop-in problem completely. Even after 12 months, patients in the experimental group still averaged slightly over one drop-in visit during a 6-month period of time. Some of these drop-in visits came from the small proportion of the sample who did not use the telephone at any time during the study. The fact that a large proportion of the calls could be handled by nonphysicians suggests that a telephone may be an effective and efficient means of providing service, so that physician time can be used more effectively. Problems encountered in scheduling outpatient visits are not unique to our center; therefore, a telephone service may have wide application not only for drop-in patients but for other problem areas in delivering high quality ambulatory care.

References

1. Andersen R. A behavioral model of families' use of health services. Chicago: University of Chicago, Center for Health Administration Studies, Research Series 25, 1968.

2. Bice TW, White KL. Factors related to the use of health services: an international comparative study. Med Care 1969;7:124.

3. Avnet HD. Physician service patterns and illness rates. New York: Group Health Insurance, Inc, 1967.

4. Greenlick MR, Freeborn DK, Colombo TJ, Prussin JA, Saward EW. Comparing the use of medical care services by a medically indigent and a general membership population in comprehensive prepaid group practice program. Med Care 1972;10:187.

5. Bice TW, Eichhorn R, Fox PD. Socioeconomic status and use of physician services: a reconsideration. Med Care 1972;10:261.

6. Mechanic D. Illness and cure. In: Dosa J, Antonavsky A, Zola IK, eds. Poverty and health. Cambridge: Harvard University Press, 1969:191.

7. Montiero L. Expense is no object . . . : income and physician visits reconsidered. J Health Soc Behav 1973;14:99.

8. Weinerman ER. Patients' perceptions of group medical care. Am J Public Health 1964;54:880.

9. Stoeclke JI, Zola IK, Davidson GE. On going to see the doctor: the contribution of the patient to the decision to seek medical aid. J Chronic Dis 1963;16:975.

10. Coe RM, Wessen AF. Social-psychological factors influencing the use of community health resources. Am J Public Health 1965;55:1024.

11. McKinlay JB. Some approaches and problems in the study of the use of services: an overview. J Health Soc Behav 1972;13:115.

12. Lewis CE. Acceptance of physicians' assistants. Hospitals 1971;45:62.

13. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966;44:166.

14. U.S. Department of Health, Education, and Welfare, Volume of Physician visits, 1966, 1967, 1969 and 1971, Public Health Service Publication No. 49, 75 Series 10, Number 49. Washington: U.S. Government Printing Office.

15. Patterson P, Bergman A, Wedgewood R. Parent reaction to the concept of pediatric assistants. Pediatrics 1969;44:69.

16. Brown SB, Eberle BJ. Use of telephone by pediatric house staff—a technique for pediatric care not taught. J Pediatr 1974;84:117.

17. Ott JE, Bellaire J, Machota P, et al. Patient management by telephone by child health associates and pediatric house officers. J Med Educ 1974;49:596.

18. Strain J, Miller J. The preparation, utilization and evaluation of a registered nurse trained to give telephone advice in a private pediatric office. Pediatrics 1971;47:1051.

19. Wang MK, Kirkham FT. Use the telephone in a health maintenance service for the chronically ill—a preliminary report. J Chronic Dis 1971;24:489.

20. Williamson JW, Goldberg E, Packard M. Use of simulated patients in evaluating patient management skills of telephone counselors, a proposal. Crisis Intervention and Counseling by Telephone.

21. Littman RE, Farberow NL, Shneidman ES, et al. Suicide-prevention telephone service. JAMA 1965;192:107.

22. Mapes RW, Lewis EW, Covell DG, et al. Feasibility study of pediatric telephone consultation service. Pediatrics 1972;50:307.

23. Grebin B, Asnes R. Dial-a-culture—a new application of the telephone to increase compliance. Abstracts of the Ambulatory Pediatric Association 1975;159.

24. Lattimer A. Telephone triage ("Totline") by nurses in a pediatric acute care clinic. Abstracts of the Ambulatory Pediatric Association 1975;188.

25. Hulka BS, Zyzanski SJ, Cassel JC, Thompson SJ. Scale for the measurement of attitudes toward physicians and primary medical care. Med Care 1970;8:429.

26. Bancroft TA. Topics in intermediate statistical methods. Ames, Iowa: Iowa State University Press, 1968.

27. Chassin MR. The containment of hospital costs: a strategic assessment. Med Care 1978;16(Suppl).

MEDICAL CARE
July 1982, Vol. XX, No. 7

Book Reviews

Antitrust and the Health Care Provider, by Martin J. Thompson, Germantown, Maryland, Aspen Systems Corporation, 1979, 250 pp., $30.00 cloth.

APPLICATION OF ANTITRUST LAW to the activities of health care providers has sent shock waves throughout the industry. Professional associations have been particularly reluctant to accept the fact that they are subject to the strictures on anticompetitive practices laid down in the Sherman, Clayton and Federal Trade Commission Acts. Martin Thompson's book comes at a time when health care providers are concerned about the antitrust implications of their collective actions. Unfortunately, the book is written primarily for lawyers and may not be geared to the audience most in need of developing a better understanding of what antitrust enforcement means in this market.

The book is divided into nine chapters, each of which, according to the author, was written to stand on its own. With the exception of the introductory chapter—a compact review of the antitrust statutes, jurisdictional issues and general implications of the antitrust laws for the health care market—each chapter discusses a discrete set of activities. The chapters vary considerably in the depth and quality of the underlying research, although they are consistently weak on including references to economic and social science research findings. Some chapters are good summaries of the statutory and case law, but overall the book lacks analysis of the policy implications of current law for health care providers. Consequently, it reads more like a hornbook for the fledgling attorney than a general treatise.

Part of the problem with this book is that it deals with an area of the law that is developing rapidly. It was written almost three years ago, and many of the legal issues covered in the book have now been litigated and decided by the courts. In fairness to the author, the merits of the book should not be judged solely on how well it anticipated subsequent events but instead on its intrinsic value.

Certain chapters of this book are well worth reading. In particular, chapter 3 ("Health Planning and other Government Interventions"), chapter 5 ("The Role of Trade Associations") and chapter 7 ("Relations with Health Insurers") represent decent summaries of the legal implications of the antitrust laws in these areas of provider activity. In chapter 3, the author reviews First Amendment protection for legitimate efforts by health care providers to influence governmental policy decisions, and other immunities from antitrust suits conveyed by the implied repeal and state action doctrines. Since these immunities are important to the disposition of antitrust cases challenging certain activities, Thompson's discussion of them this early in his book was a wise choice.

Chapter 5 is perhaps the best 40 pages in the book. Thompson briefly runs through the role of professional associations in setting standards through accreditation and certification, organizing group boycotts of nonmedical providers, conducting disciplinary proceedings, conducting boycotts of insurers and HMOs, setting fees, organizing peer review activities and lobbying. The only jarring inconsistency in this chapter is the inclusion of cooperative buying and selling arrangements in the sweep of professional association conduct. These activities deserve far more consideration than the two-page discussion here; ideally, they belong in a separate chapter along with similar arrangements used by hospitals or other health care institutions.

Chapter 7 is a concise little chapter that deals with provider/insurer relations—a topic of considerable importance in this industry. Thompson begins with a detailed analysis of the McCarran–Ferguson Act (which exempts the "business of insurance" from antitrust scrutiny), and concludes with a review of insurance boycotts, coercion and intimidation. It is in this latter area that antitrust enforcement is likely to intensify in the future as insurers attempt to hold the line on costs with providers. In fact, the Federal Trade Commission and the Department of Justice have brought several cases against providers who allegedly organized such boycotts. But the author does not share his views on where the case law appears to be going; instead, he merely recites current law. The result is a wooden cross-sectional look at the status quo, with no consideration of what that look portends for the future.

The remaining chapters are pedestrian restatements of the antitrust laws. Chapter 4, "Mergers, Acquisitions, and Affiliations," reviews an area of the law that, until recently, has had little application to health care institutions. Only since the rapid growth of hospital and

C5170-1

Medical Care

August 1982
Vol. 20 No. 8
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

A Review of Methods for Ambulatory Medical Care Evaluations

An Examination of the Factors That Influence Patient Referral Decisions

Use of Illness-Related Ambulatory Physician Services in Finland

Visits to Physicians by Asian/Pacific Americans

Clinical Evaluation vs. Economic Evaluation: The Case of a New Drug

Drug Product Selection in Illinois: The First Year's Experience

Evaluation of a Sequence of Diagnostic Tests Using the Workup of Ureteral Stone as a Model

The Use of Ancillary Services: A Role for Utilization Review?

Podiatric Surgery: The Need for a Second Opinion

Utilization of an Automated Multiphasic Health Testing System for Performing Prehospitalization Examinations

PUBLISHED BY

Medical Care
August 1982, Vol. XX, No. 8

An Examination of the Factors That Influence Patient Referral Decisions

Robert L. Ludke, Ph.D.

The results of a study to identify the factors that two types of referring physicians, general/family practitioners and general surgeons, consider when deciding whether or not to refer the patient, and where to refer the patient are presented and discussed. The factors for each type of decision were identified initially through interviews with physicians in active practice in Wisconsin and rank ordered according to their relative importance in making the referral decisions. These factors were then used in two alternative normative decision-making models to predict physician referral behavior. Factors related to the kind and quality of medical care the patient would receive appear to be most important in the decisions, with a high level of agreement between and among general/family practitioners and general surgeons as to the relative importance of the factors. Although the normative decision-making models correctly predicted a relatively high percentage of the referral decisions, not all of the factors were found to be important in explaining the behavior of physicians, and were not always consistent with the factors that were reported to be important.

AT SOME POINT during the process of delivering medical care to a patient, a physician may consider referring that patient to another physician or medical center. The decision to refer a patient may occur at any time during the care process, such as during diagnosis, after diagnosis but before treatment, or during treatment. The reason for referral may vary. It may be for diagnosis, confirmation of diagnosis, opinion regarding treatment, or treatment, and may depend upon the scope of practice of the referring physician, the diagnostic and treatment equipment and procedures available to the physician, time constraints, and a number of other potential factors related to the physician, the patient, and the referral physicians or centers. In contrast to consultation, where responsibility for a patient remains with the initial physician but advice or special studies from other physicians are requested, referral represents that act involved in transferring some or all responsibility for the patient's care, either temporarily or permanently, from one physician or health agency to another physician or health agency.[1-3]

Studies conducted in the United States found that physicians refer approximately 1–6 per cent of their patient visits, on the average, with large variation in referral rates among individual physicians.[1-11] Several concerns have been raised regarding the referral process. Clute,[12] Miller,[13] and Peterson et al.[14] have questioned whether appropriate referrals are being made. Others have raised questions about the adequacy of communication between

---

Graduate Program in Hospital and Health Administration, College of Medicine, The University of Iowa.

Supported by Grant Number 1 R03 HS 02351-0 from the National Center for Health Services Research, HRA.

Address for reprints: Robert L. Ludke, Ph.D., Graduate Program in Hospital and Health Administration, S-502 Westlawn, The University of Iowa, Iowa City, IA 52242.

0025-7079/82/0800/0782/$01.55 © J. B. Lippincott Co.

referring and consulting physicians.[1,6,11,15–17] Williams et al.[18] found failures on the part of consulting physicians to return patients to referring physicians.

Although previous research has provided some insight into the rates of referral relative to physician and patient demographic characteristics and into some of the major problems that appear to exist in the referral process, it contributes little to an understanding of why and how the process operates as it does. The study conducted by Shortell[3] is an exception, however. His model of physician referral behavior was based on social exchange theory. The basic premise was that physicians, occupying different levels of status within the medical community, perceive different rewards and costs associated with referring patients. Shortell concluded,

> In summary, between seven and 49 per cent of the variance in the referral rates is explained by exchange theory and related variables . . . neither the status index nor individual components of the index figures significantly in explaining the results . . . other variables such as board certification, organization of practice, caseload severity, and office visits per week explain much of the variance not accounted for by the status variable. (p. 69)

A subsequent study by Shortell and Vahovich[8] was conducted to address the hypothesis that patient-related variables are the most reliable predictors of referral rates for client-dependent physicians, such as general practitioners, and physician-related variables are the most reliable predictors of referral rates for colleague-dependent specialists, such as surgeons. Although they found that physician-related variables were the most important predictors of referral rates for surgeons, they also found that to be the case for general practitioners.

Given the frequency of referrals, the implications of the referral decisions and the potential concerns regarding the referral process, the referral process warrants further study. Particularly, greater insight into the variables which physicians take under consideration in making the referral decisions is needed. Through a greater understanding of the referral process, steps can be taken to begin assessing whether or not optimal use is being made of referring physicians, the referral process, and referral centers in terms of manpower, resources, and quality and cost of patient care. Also, insights can be gained into how and what modifications should and could be made in order to strengthen the referral system.

This paper takes a step in that direction by presenting and discussing the results of a study to identify the factors that physicians consider when faced with the possibility of referring a patient. The study focused on the referral of breast cancer patients after the referring physician conducted the initial history and physical examinations of the patient and performed such preliminary diagnostic procedures as chest x-ray, liver function tests, blood count and/or mammogram. Given this point in time in the patient's care process, the physician must decide whether or not to refer the patient for further diagnostic workup and/or treatment and if the decision is to refer the patient, where to refer that patient.

Since the factors that are considered may differ for each of these decisions, the study assessed the factors important for each decision. In addition, the study looked at two types of physicians, general/family practitioners and general surgeons, to determine if the factors differ across types of physicians. These types of physicians were selected because they serve as a major source of referrals for breast cancer patients in Wisconsin, where the study was conducted.

In 1979, Wisconsin had the 16th highest incidence rate of new breast cancer cases per 100,000 population and the 11th highest breast cancer mortality rate per 100,000 population in the nation.[19] There were approximately 1,015 licensed general/family practitioners and 418 licensed general surgeons in office-based practice in the state at the time the study was conducted.[20] These physicians had a number of centers available, located either in Wisconsin or adjacent states and associated with either university medical schools, such as Wisconsin, Minnesota, or Chicago, or large multispecialty group practices, such as the Mayo Clinic and the Marshfield Clinic, for the referral of their patients. In addition, there were a number of oncologists associated with large hospitals in the major cities within or bordering the state which also could be used for patient referral.

Methods

The study consisted of three phases. The first phase identified the factors that appeared to be important in deciding whether or not to refer a patient and where to refer that patient. Given the lack of understanding of the referral process from the literature, two field studies were conducted to gain insight into how, where, when and why patients are referred. In the first study, the author informally interviewed 38 physicians in active practice in Wisconsin, the majority of whom were primary care physicians.[21] The physicians were not selected using any sampling scheme but were physicians who were thought to be open and informative about the referral process and who would be willing to talk to the author about it for an hour. The interviews were content analyzed to identify the factors which the physicians reported as influencing their referral decisions. Because patients often were reported as playing a role in the referral decisions, the author, in the second field study, interviewed 23 patients who had been referred to a major medical center to clarify this role.[22] The results of these studies, as well as the pertinent referral literature, served as the basis for identifying the referral factors addressed in this study.

During the second phase of the study, 11 general surgeons and 14 general practitioners in active practice in Wisconsin rank ordered the relative importance of the factors in making the referral decisions. These physicians either had participated in the physician field study or were nominated by physicians who had participated in the study. Each physician was first given a deck of 3 × 5 cards with one refer/not refer factor and accompanying definition on each card. After reading through all of the cards, the physician rank ordered the factors according to the perceived importance in deciding whether or not to refer a patient. Second, each physician was given a deck of 3 × 5 cards listing the where-to-refer factors and accompanying definitions, and asked to rank order these factors in a similar manner. This task was done on an individual physician basis in the physician's office with no interaction among physicians.

The third phase of the study attempted to predict referral decisions using the identified factors. Two alternative decision making models, which have been proposed and researched extensively as normative models to facilitate the decision making process, were used.[23–25] These models provided alternative ways of looking at a decision-making situation, either through probabilities or utilities, and have been found in some situations to outperform regression models in predicting individual decision-making behavior.[25,31] Also, the use of these models overcome the problem of not having an actuarial data base from which to develop regression models.

The first model was the Multi-Attribute Utility (MAU) model, which in its simple linear form can be expressed as:

$$U_k = W_1U_1(F_1) + W_2U_2(F_2) + \cdots + W_NU_N(F_N)$$

where U_k = the utility or preference associated with the k *th* decision, such as refer or not refer, F_i = the level of the *ith* factor present with the k*th* decision ($i = 1,2, \cdots ,N$), $U_i(F_i)$ = the utility or preference associated with the level of the i *th* factor (i = 1,2, $\cdots$,N), and W_i = a scaling constant that reflects the relative importance of the i *th* factor on the overall utility of the k*th* decision, such that $0 \leq W_i \leq 1$ and $\Sigma W_i = 1$ (i = 1,2, $\cdots$,N).

Thus, given the relative importance of each of the referral factors in making a referral decision, such as the decision to refer/not refer, the levels of each of the referral factors present for each alternative decision and the relative preference of each of those factor levels, the MAU model calculates a value representing the relative preference or utility of each alternative decision. The decision with the greatest utility should be the decision that is made. This model has been found to be a valid predictor of decisions regarding health program funding,[26] job search and applicant placement,[27,28] degree of burn severity[29] and physician shortage areas.[30]

The second model was the conditionally independent ratio form of Bayes' Theorem referred to as the Bayesian Probability (BP) model, which can be expressed as:

$$\frac{P(D_1|F_1,F_2,\cdots,F_N)}{P(D_2|F_1,F_2,\cdots,F_N)} = \frac{P(F_1|D_1)}{P(F_1|D_2)} \cdot \frac{P(F_2|D_1)}{P(F_2|D_2)} \cdots\cdots \frac{P(F_N|D_1)}{P(F_N|D_2)} \cdot \frac{P(D_1)}{P(D_2)}$$

where

$\dfrac{P(D_1|F_1,F_2,\cdots,F_N)}{P(D_2|F_1,F_2,\cdots,F_N)}$ = the posterior odds of decision D_1 being made in comparison to decision D_2 given the levels of factors F_1, F_2, $\cdots$,F_N are present, $P(F_i|D_1)$ = the likelihood ratio of the level of factor F_i being present given decision D_1 is made in comparison to decision D_2 (i = 1,2, $\cdots$,N), $P(D_1)$ = the prior odds of decision D_1 being made in comparison to decision D_2 without the levels of factors $F_1,F_2, \cdots,F_N$ being present.

Thus, given the prior odds of one alternative decision being made compared with another, such as refer the patient versus not refer the patient, and the likelihood of each of the factor levels being present if one alternative decision is made compared with another, the BP model calculates the odds of one decision being made compared with the other. From this posterior odds, the probability of each of the decisions being made can be calculated by:

$$P(D_1|F_1,F_2,\cdots,F_N) = \frac{\text{Posterior Odds}}{1 + \text{Posterior Odds}}$$

$$P(D_2|F_1,F_2,\cdots,F_N) = 1 - P(D_1|F_1,F_2,\cdots,F_N)$$

The decision with the largest probability is the decision that should be made. Models of this form have been used in the decision making areas of medical diagnosis,[31–33] intelligence gathering,[34,35] job selection[36] and weather forecasting.[37]

Because of the lack of actuarial data regarding the referral factors, the likelihood ratios of the BP model and the factor weights and utilities of the MAU model were subjectively estimated by eight general surgeons (GSs) and eight general practitioners (GPs) in active practice in Wisconsin. These physicians were selected randomly from the physicians who rank ordered the referral factors in the second phase of the study. The remaining physicians were used to pilot test the estimation procedures and, thus, their results were not included in this phase of the study. Eight physicians, four GSs and four GPs, made estimates for the MAU model and eight made estimates for the BP model. Determination of which physicians estimated which parameters was done on a random basis.

For the BP model, each physician was asked to consider a situation where 100 physicians who referred their last breast cancer patient to a consultant were selected randomly. Given this was all the information about these physicians, the patients, and the consultants, the physician estimated how many of these 100 physicians would perceive, for example, that 1) they provide a higher quality of patient management than a consultant, 2) they provide the same quality of patient management as a consultant, and 3) a consultant provides a higher quality of patient management than they. After completing this task, each physician repeated the approach for the situation where 100 physicians did not refer their last breast cancer patient. This task was completed for each of the refer/not refer factors. The likelihood ratio for each factor level was calculated by dividing the likelihood estimate given the decision to refer, by the estimate given the decision not to refer. A similar approach was used to make likelihood estimates for the where-to-refer factors.

For the MAU model, the estimation task was divided into two phases. First, after ranking the relative importance of the refer/not refer factors, each physician rated the relative importance of each factor, assigning the least important factor an arbitrary value of "10" and then giving every other factor a score reflecting its relative importance. For example, if a physician thought a factor was 1½ times as important as the least important factor, a score of "15" would be assigned. No upper bounds were placed on the scores and values reflecting equal importance were permitted. Individual physician factor weights then were calculated by normalizing these ratings. A similar approach was used to estimate the factor weights for the where-to-refer factors.

The second phase required each physician to estimate the utilities of the MAU model by assigning preference scores to each level of the refer/not refer and where-to-refer factors. Each physician was given a preference sheet for each factor, which listed five factor levels. The physician assigned a score of 0 to the least preferred level, a score of 100 to the most preferred level and scores in between reflecting relative preferences for every other level.

Estimates for both models were made on an individual physician basis without interaction. The averages of the estimates across physicians were used in the models.

In addition, 10 GPs and five GSs in active practice in nonmetropolitan areas in Wisconsin each reviewed four hypothetical case descriptions of breast cancer patients, which were based on actual patients. These physicians had not participated in the earlier phases of the study and were selected on the basis of their willingness to provide two hours of time to evaluate the cases. To test the ability of the models to predict referral as well as nonreferral decisions, the cases were designed, with the assistance of a fellow of the University of Wisconsin's Clinical Cancer Center, to be representative of breast cancer patients as they might present themselves in a primary care physician's office, ranging in severity from those that a general practitioner or general surgeon would tend not to refer to those that they would tend to refer immediately. After reading a case description, the physician was asked whether or not he/she would refer the patient and, if so, where. The physician was not provided with a list of alternative referral centers, but indicated his/her first and second choices for consultants. In addition, after reading the definitions of the referral factors, each physician specified the level of each factor involved in making the decisions. This task was completed on an individual physician basis, without interaction, in each physician's practice setting.

The information obtained from the referring physicians on the levels of the factors involved in making the decisions was then used by the models to calculate a predicted decision. This predicted decision was compared with the actual decision made by the referring physician. The per cent of decisions correctly predicted by each of the models was then determined. These percentages were analyzed using a split-plot repeated measures design, with one between-block treatment (type of physician making the referral decisions) and one within-block treatment (type of model).[38] Because the per cent scores were calculated on different sample sizes (five GSs vs. 10 GPS), weighted scores were used in the analyses to insure equal cell variances. A 0.05 significance level was used in all analyses.

Results

The field studies indicated that 12 factors appeared to be important in deciding whether or not to refer a patient and 15 factors appeared to be important in deciding where to refer a patient. The Appendix presents a brief definition of each factor. As illustrated in Tables 1 and 2, several factors appear under both decisions because these factors were perceived by physicians as important issues, regardless of whether the patient was treated by the referring physician or a consultant, or whether the patient was treated by one of several possible consultants. Also, the factors appear to be of three basic types. First are those factors, such as "patient results," which relate to the treatment/care the patient would receive. Second are those noncare related factors, such as "convenience to the patient," which are associated with the patient and/or family. Third are noncare related factors, such as "reciprocations physician receives," which are associated with the referring physician.

TABLE 1. Decision to Refer/Not Refer Factors

Technical/care related factors
Quality of patient management
Patient results
Individualized patient management and care
Patient related factors
Convenience to the patient
Cost to the patient
Patient's potential for productive remaining life
Patient expectations for referral
Physician related factors
Physician's loss of patient/family/income
Personal benefits physician receives
Physician's satisfaction with previous referrals
Referring physician's degree of belief referring is admitting failure
Attitude of physician's colleagues/local medical community towards referring

Table 3 presents the rank orderings of the 11 GSs and the 14 GPs for the decision to refer/not refer factors, along with the average GS and GP rank orderings. There is significant agreement among both the GSs and the GPs as to the relative importance of the factors, with slightly greater agreement among the GPs than the GSs. The Spearman rank correlation coefficient calculated between the average GS and

TABLE 2. Decision Where to Refer Factors

Technical/care related factors
Quality of patient management
Patient results
Individualized patient management and care
Patient related factors
Convenience to the patient
Cost to the patient
Degree patient preferences are met
Previous use of and satisfaction with consultant by patient/family
Physician related factors
Physician's loss of patient/family/income
Reciprocations physician receives
Communication physician receives
Respect/courtesy physician receives
Attitude of patients towards consultant
Physician's personal knowledge of consultant
Physician's satisfaction with previous referrals to consultant
Attitude of physician's colleagues/local medical community towards consultant

average GP rank orderings indicates a significantly high level of agreement between the two types of physicians. The most important factors for both types of physicians were the technical/care related factors of "quality of patient management," "patient results" and "individualized patient management and care," followed by "patient's potential for a productive remaining life." The least important factors for both physicians were the physician-related factors of "referring physician's degree of belief referring is admitting failure," "physician's loss of patient/family/income" and "attitude of physician's colleagues/local medical community towards referring."

Similarly, Table 4 presents the rank orderings for the decision where-to-refer factors. Inspection of Table 4 indicates again significant agreement both among GSs and GPs as well as between GSs and GPs as to the relative importance of the factors. Again, those factors related to the treatment aspects of a patient's care were ranked most important as well as the "physician's personal knowledge of the consultant." The least important factors were the physician-related factors of "physician's loss of patient/family/income," "attitude of physician's colleagues/local medical community towards consultant," "cost to the patient," "reciprocation physician receives," and "respect/courtesy physician receives."

Table 5 indicates that, on the average, the MAU and BP models using all of the referral factors were able to correctly predict 74.2 per cent of the decisions to refer/not refer and 81.0 per cent of the decisions of where to refer. The analyses of variance revealed highly significant differences between the models for the decisions to refer/not refer, with the BP model correctly predicting a significantly greater per cent of the GS decisions than the GP decisions (90.0 per cent vs. 70.0 per cent), while the MAU model correctly predicted a significantly greater per cent of the GP decisions than the GS decisions (77.5 per cent vs. 60.0 per cent). For the decision where to refer, no significant differences between models were found. On the average, the models correctly predicted a greater per cent of GS decisions than GP decisions (91.7 per cent vs. 76.7 per cent).

However, models using fewer referral factors were found to predict better than or as well as the models using all of the referral factors (Table 6). For the decision to refer/not refer, a three factor MAU model (quality of patient management, patient results, and attitude of physician's colleagues/local medical community towards referring) and a four factor BP model (quality of patient management, patient expectations for referral, referring physician's degree of belief referring is admitting failure and attitude of physician's colleagues/local medical community towards referring) were found to predict a greater per cent of the decisions correctly than the respective total factor models. No significant differences were found between these fewer factor models nor were differences found in the performances of the models between GS and GP decisions.

What is of interest are the factors included within these models. Both models included the factor "quality of patient management," which was ranked as the most important factor by both the GSs and GPs and the factor "attitude of physician's colleagues/local medical community towards referring," which was ranked third least important by the GS and GP physicians. In addition, the BP model included the factor "referring physician's degree of belief referring is admitting failure," which was ranked least important by the physicians.

To further address this issue, a stepwise regression analysis was performed to identify the set of factors that best explained the decisions, using the ratings of the factors provided by the referring physicians as the independent variables and the referring physicians' decisions (refer vs. not refer) as the dependent variables. "Attitude of

TABLE 3. General Surgeon (GS) and General Practitioner (GP) Rankings of the Decision to Refer/Not Refer Factors (12—Most Important, 1—Least Important)

Factors	General Surgeons												General Practitioners														
	1	2	3	4	5	6	7	8	9	10	11	Average	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Average
Quality of patient management	12	9	12	8	12	6	11	12	10	12	12	12	12	11	12	10	11	10	11	12	12	12	12	10	12	10	12
Patient results	11	11	6	5	10	11	8	9	9	11	11	11	4	12	9	12	12	9	10	9	11	10	2	9	11	8	10
Individualized patient management and care	9	10	3	7	9	5	12	11	11	10	10	10	11	10	10	7	10	11	8	11	10	11	11	7	10	9	11
Convenience to the patient	6	5	11	11	8	3	6	5	8	8	7	8	9	6	6	3	6	8	6	10	9	5	7	5	6	6	6
Cost to the patient	5	6	5	10	7	9	4	4	7	7	6	5	5	3	7	5	7	7	7	8	8	4	5	6	7	5	5
Physician's loss of patient/family income	1	1	1	1	2	4	3	3	1	6	1	2	3	2	2	4	3	1	1	1	2	1	3	3	3	2	2
Personal benefits physician receives	7	4	9	4	5	7	7	7	3	5	5	4	8	7	5	6	2	2	5	4	4	8	6	4	4	4	4
Patient's potential for productive remaining life	10	12	4	3	11	8	5	8	12	9	9	9	10	8	11	11	8	3	12	6	7	9	9	11	8	12	9
Patient expectations for referral	4	3	7	12	4	12	10	6	6	1	8	6	7	9	8	8	5	12	4	7	6	6	8	12	9	7	8
Physician's satisfaction with previous referrals	8	8	10	6	6	10	9	10	4	2	4	7	6	5	3	9	9	6	9	5	5	7	10	8	5	11	7
Referring physician's degree of belief referring is admitting failure	3	2	2	2	1	2	2	2	2	3	2	1	1	1	1	1	1	4	2	2	1	2	1	1	1	3	1
Attitude of physician's colleagues/local medical community towards referring	2	7	8	9	3	1	1	1	5	4	3	3	2	4	4	2	4	5	3	3	3	3	4	2	2	1	3
Kendall Coefficient of concordance												0.5464 (p<.001)															0.7328 (p<.001)

Spearman Rank Correlation Coefficient between average GS and GP rankings = 0.9650 ($p < .01$).

TABLE 4. Continued.

Physician's personal knowledge of consultant	12	11	12	8	11	12	13	12	12	12	6	13	11	13	12	11	11	15	12	9	8	12	6	7	12	7	12
Physician's satisfaction with previous referrals to consultant	8	12	14	9	10	11	10	11	11	11	7	11	10	9	6	9	12	13	11	5	7	13	7	6	14	11	10
Attitude of physician's colleagues/local medical community towards consultant	2	3	8	7	8	1	2	1	5	1	4	2	1	2	4	1	8	3	2	3	3	2	1	2	2	4	2
Kendall Coefficient of Concordance												0.5821 (p<.001)															0.6368 (p<.001)

Spearman Rank Correlation Coefficient between average GS and GP rankings = 0.9759 ($p < .01$).

TABLE 5. Mean Per Cent Correct for the Total Factor MAU and BP Models by Type of Referring Physician for the Decisions to Refer/Not Refer and Where to Refer

Decision	Referring Physician	MAU Model	BP Model	Average
Refer/not refer	GS	60.0	90.0	75.0
	GP	77.5	70.0	73.8
	Total	71.7	76.7	74.2
Where to refer	GS	91.7	91.7	91.7
	GP	76.7	76.7	76.7
	Total	81.0	81.0	81.0

physician's colleagues/local medical community towards referring" was found to be the best explainer variable, followed by "patient expectations for referral" and "quality of patient management." These three factors explained approximately 41 per cent of the variance in the decisions. Two of these factors "attitude of physician's colleagues/local medical community towards referring" and "quality of patient management" were included in the three factor MAU model and all three factors were included in the four-factor BP model.

Similarily, for the decision of where to refer, a five-factor MAU model (degree patient's preferences are met, reciprocations physician receives, communication physician receives, previous use of and satisfaction with consultant by patient/family, and physician's personal knowledge of consultant) and a four-factor BP model (same factors as the five factor MAU model, except communication physician receives) were found to correctly predict a greater per cent of the decisions than the respective total factor models (Table 6). No significant differences were found between the fewer factor models nor were differences found in the performances of the models between GS and GP decisions.

Interestingly, the technical/care-related factors were absent from the fewer factor models for the decision of where to refer. An explanation for this may be related to

TABLE 6. Mean Per Cent Correct for the Fewer Factor MAU and BP Models by Type of Referring Physician for the Decisions to Refer/Not Refer and Where to Refer

Decision	Referring Physician	MAU Model	BP Model	Average
Refer/not refer	GS	85.0	85.0	85.0
	GP	85.0	90.0	87.5
	Total	85.0	88.3	86.6
Where to refer	GS	91.7	83.3	87.5
	GP	80.0	83.3	81.6
	Total	87.5	83.3	85.4

one of the findings of the physician field study. During that study, physicians indicated that the treatment/care-related factors were the most important factors and would be considered first in making the referral decisions. If those factors were equivalent for various consultants, then the other factors would play a major role in the decisions. If this is the case, and the consultants, who were specified by the physicians when they indicated where they would refer patients, were perceived as equivalent in terms of the technical/care-related factors, then these factors should be absent from the models. Lack of variability of the dependent variable, type of decision, prohibited a stepwise regression analysis for the decision where to refer.

Conclusions

The results of the study have indicated that there are a number of factors, most of which have not been studied in the referral literature, that appear to be important when a physician is faced with the possibility of referring a patient. Of prime importance are factors related to the kind and quality of medical care the patient would receive. Although there seem to be some individual differences, generally there appears to be a high level of agreement between and among different types of referring physicians as to the relative importance of the factors. Also, there did not seem to be any additional support for Shortell and Vahovich's hypothesis that patient-related factors may be more important to client-dependent physicians, such as general practitioners, and physician-related factors may be more important to colleague-dependent specialists, such as surgeons.

The normative decision-making models correctly predicted a relatively high per cent of referral decisions. If the models, whose parameters were subjectively estimated, have some internal validity, these results lend some credibility to the identified factors. However, it also was found that not all of the factors appear to be important in predicting the referral decisions. The factors found to be the best predictors, through the models and a stepwise regression analysis, were not always consistent with the factors reported by the physicians to be the most important factors. These results raise questions about differences between what physicians may report to be important, what they may actually think is important and how they behave, which must be investigated further.

It is interesting to note the apparent importance of the factor "attitude of physician's colleagues/local medical community towards referring" in predicting the decisions to refer/not refer. As suggested by physicians interviewed during the field studies, a physician's colleagues may have a significant impact on his/her referral behavior to the point of determining which patients will be referred. Because of the implications such norms may have, this issue needs further investigation.

The results of this study suggest major implications for consulting physicians or centers. Programs for accepting patient referrals, diagnosing and/or treating those patients and returning the patients to the referring physicians, should be structured and operated in such a way that recognize the factors that a consultant's potential clientele consider important in the referral process. The results of this study imply that it is imperative for a consultant to first have the technical capability and expertise to provide high quality patient management with positive patient results, in order to maintain a viable referral network. However, the consultant must then address the other physician- and patient-related factors of importance to the referring physicians. Consultants should attempt to personalize their relations with referring physicians. The results of the study indicate that personal knowledge of the consultant appears to be an important consideration when referring physicians decide where to refer patients. This implies that consultants should take some initiatives in overcoming the town-and-gown syndrome by, for example, visiting potential referring physician clientele and helping those physicians better address their problems. Some referring physicians expect consultants to reciprocate patient referrals through education on the diagnoses and/or treatment of various types of cases, training of support staff personnel and other ways of improving his/her capabilities. The study also indicated that the referring physician's colleagues and/or local medical community play a role in the referral process, implying the need for consultants to educate the medical community regarding the diagnosis and treatment of various types of problems, the appropriate use of referrals and their capabilities and expertise. This could be accomplished, for example, through continuing education programs jointly sponsored by the consultants and local medical societies.

The results of the study also indicate that the patient does play a role in the referral process. Referring physicians appear to be attuned to the patient's expectations for referral, preferences for certain consultants or referral centers, and previous use of and satisfaction with consulting physicians or centers. This implies that consultants must be aware of this role, realizing that the patient provides an important feedback mechanism not only back to the referring physician but also to friends and neighbors. The consultants need to insure that services are provided to patients in such a way that every patient is a potential "sales representative" of their services. In addition, the study implies that consultants should educate the patient in appropriate referral behavior to insure that the patient does not place inappropriate referral demands on referring physicians.

This study found that little actually is known about the referral process in terms of how it operates; the relationships between referring physicians, patients and consultants; the decision making processes of each of the key participants and the environmental context which influences the operation of the process. Research is needed to understand better the transactions between the referring physician and the patient regarding the referral decisions, and the transactions between the referring and consulting physicians during a referral encounter. Although this study dealt with the referral of breast cancer patients, the applicability of the results and conclusions may extend beyond this limited sample of referrals to the general population of patient referrals. The results of the field studies indicated that the referral factors addressed in this study are applicable to any type of patient referral.[21] However, what is not known is the role these factors play in different types of referral decisions. For example, the relative importance of the factors may change for different types of referral patients. Further research must be conducted to further validate the findings of this study in order that the results may be generalized to other types of patients and referral transactions. Also, research is required on developing measures of effectiveness, not only in terms of process but also outcome, that can be used to assess and evaluate the referral process and to make appropriate modifications to maximize a referral program's efficiency and effectiveness.

References

1. Metcalfe DHH, Sischy D. Patterns of referral from family practice. J Fam Prac 1974;1:34-8.

2. Penchansky R, Fox D. Frequency of referral and

patient characteristics in group practice. Med Care, 1970;8:368-84.

3. Shortell SM. A model of physician referral behavior: a test of exchange theory in medical practice. Research series no 31. Chicago: Center for Health Administration Studies, University of Chicago, 1972.

4. Avnet HH. Physician service patterns and illness rates: a research report on medical data retrieved from insurance records. New York: Group Health Insurance, Inc, 1967.

5. Geyman JP, Brown TC, Rivers K. Referrals in family practice: A comparative study by geographic region and practice setting. J Fam Pract 1976;3:163-7.

6. Ruane TJ. Consultation and referral in a Vermont family practice: A study of utilization, specialty distribution, and outcome. J Fam Pract 1979;8:1037-40.

7. Schmidt DD. Referral patterns in an individual family practice. J Fam Pract 1977;5:401-3.

8. Shortell SM, Vahovich SG. Patient referral differences among specialities. Health Serv Res 1975;10:146-61.

9. Taubenhaus LJ. A study of one rural practice, 1953. General Practitioner 1955;12:96-102.

10. Tenney JB, White KL, Williamson JW. National ambulatory medical care survey: background and methodology, (Vital and Health Statistics, Series 2, No 61) DHEW Publication: HRA 74-1335, 1974.

11. Williams TF, White KL, Fleming WL, Greenberg BG. The referral process in medical care and the university clinic's role. J Med Educ 1961;36:899-907.

12. Clute KF. The general practitioner: a study of medical education and practice in Ontario and Nova Scotia. Toronto: University of Toronto Press, 1963.

13. Miller MH. Who receives optimal medical care? J Health Soc Behav 1973;14:176-82.

14. Peterson OL, Andrews LP, Spain RS, Greenberg BG. An analytical study of North Carolina general practice. J Med Educ 1956;31:1-165.

15. Bates BB, Torkelson LO. Two methods of improving referral information: a comparative study in a university medical center. Med Care 1967;5:418-22.

16. Cummins RO, Smith RW. Reports to referring physicians. N Engl J Med 1975;293:361, 362.

17. Kunkle EC. Communication breakdown in referral of the patient. JAMA 1964;187:663.

18. Williams TF, White KL, Andrews LP, Diamond E, Greenberg BG, Hamrick AA, et al. Patient referral to a university clinic: Patterns in a rural state. Am J Public Health 1960;50:1493-1507.

19. American Cancer Society. Cancer facts and figures, 1979. New York: American Cancer Society, 1979.

20. American Medical Association. Physician distribution and medical licensure in the US. Chicago: American Medical Association, 1977, 1979.

21. Ludke RL. Physician referrals: results of a field study of physicians. Madison: University of Wisconsin-Madison, 1977, unpublished report.

22. Ludke RL. Physician referrals: results of a field study of patients. Madison: University of Wisconsin-Madison, 1977, unpublished report.

23. Edwards W. Behavioral decision theory. Ann Rev Psychol 1962;12:473-98.

24. Slovic P, Fischoff B, Lichtenstein S. Behavioral decision theory. Ann Rev Psychol 1977;28:1-39.

25. Slovic P, Lichtenstein S. Comparison of Bayesian and regression approaches to the study of information processing judgment. Organizational Behavior and Human Performance 1971;6:649-744.

26. Stimson D. Utility measurement in public health decision making. Management Science 1969;15:17-30.

27. Fischer G. Convergent validation of decomposed multi-attribute utility assessment procedures for risky and riskless decisions. Organizational Behavior and Human Performance 1977;18:295-315.

28. Huber GP, Daneshgar R, Ford D. An empirical comparison of five utility models for prediciting job preferences. Organizational Behavior and Human Performance 1971;6:267-82.

29. Gustafson DH and Holloway D. A decision theory approach to measuring severity in illness. Health Serv Res 1975;10:97-106.

30. Health Services Research Group, Center for Health Systems Research and Analysis, University of Wisconsin. Development of the index of medical underservice. Health Serv Res 1975;10:168-80.

31. Gustafson DH. Evaluation of probabilistic information processing in medical decision making. Organizational Behavior and Human Performance 1969;4:20-34.

32. Gustafson DH, Kestly J, Ludke RL, Larson F. Probabilistic information processing: implementation and evaluation of a semi-PIP diagnostic system. Comput Biomed Res 1973;6:355-70.

33. Greist J, Gustafson DH, Stauss F, Rowse G, Laughren T, Chiles J. A computer interview for suicide-risk prediction. Am J Psychiatry 1973;130:1327-332.

34. Edwards W, Phillips L. Man as transducer for probabilities in Bayesian command and control systems. In: Bryon GL, Shelly MW, eds. Human judgments and optimality, New York: John Wiley and Sons, 1964.

35. Kelly C, Peterson C. Probability estimates and probabilistic procedures in current-intelligence analysis. IBM Rep 71-5047. Gaithersburg, Md.: IBM 1971.

36. Ford D, Huber GP, Gustafson DH. Predicting job choices with models that contain subjective probability judgments: an empirical comparison of five models. Organizational Behavior and Human Performance 1972;17:397-416.

37. Julian P, Murphy A. Probability and statistics in meteorology: a review of some recent developments. Bulletin of the American Meteorological Society 1972;53:957-63.

38. Kirk R. Experimental design: procedures for the behavioral sciences. Belmont, Calif.: Wadsworth, 1968.

Appendix

Definitions of the Referral Factors

Technical/Care-Related Factors

Quality of Patient Management. Thoroughness of workup and accuracy of staging; thoroughness of problem/treatment explanation to patient/family; thoroughness and efficiency of treatment; comprehensiveness of rehabilitation.

Patient Results. How patients look; complication rate; pre, during and post treatment mortality rates; cure rates.

Individualized Patient Management and Care. Management and care of the patient provided according to the specific needs and condition of the patient in a friendly, empathetic, personal manner.

Patient-Related Factors

Cost to the Patient. Travel expenses incurred by the patient/family in travelling to receive care including food and lodging; other expenses involved, e.g., babysitting costs, lost wages, etc.; out-of-pocket expenses for care.

Convenience to the Patient. Distance; degree of physical discomfort involved in travelling; ease in locating where to go and parking; availability of suitable accommodations nearby; extent of disruption of patient's/family's home life.

Patient's Potential for Productive Remaining Life. Patient's life expectancy and quality of remaining life.

Patient Expectations for Referral. Patient's and/or family's implicit or explicit desire to be referred or not referred.

Degree Patient's Preferences are Met. Expressed patient preferences regarding consultant; convenience such as distance or ability to stay with family or friends, costs.

Previous Use of and Satisfaction with Consultant by Patient/Family. Patient/family having used consultant for same, related, or nonrelated problems and their level of satisfaction with consultant.

Physician-Related Factors

Physician's Loss of Patient/Family Income. Patient/Family not returning to physician for care; consultant performing procedures which physician could perform.

Personal Benefits Physician Receives. Help with problem; education about problem diagnosis and treatment; reciprocations; avoiding hassles of referring.

Physician's Satisfaction with Previous Referrals. Satisfaction with results; the way patients were treated; the way the physican was treated.

Referring Physician's Degree of Belief Referring is Admitting Failure. Extent to which the physician believes that referring a patient is openly admitting he/she is no longer able to care for that patient.

Attitude of Physician's Colleagues/Local Medical Community Towards Referring. Norms that the physician's colleagues and local medical community have established for referring or not referring type of patient.

Reciprocations Physician Receives. What the physician may gain by referring patients to the consultant such as education, consultant coming to hold special clinics to train personnel in special procedures, patient referrals, etc.

Communication Physician Receives. Detail and quality of the reports; the promptness of the reports; admission and discharge notices; and notices of crises that arise.

Respect/Courtesy Physician Receives. Consultant's attitude towards helping and towards the physician in front of patients and others.

Attitude of Patients Toward Consultant. How satisfied/dissatisfied physician's patients previously referred to consultant were with that consultant.

Physician's Personal Knowledge of Consultant. Having worked or trained with consultant or know personally through some other means such as previous referrals, meetings, conferences, etc.

Attitude of Physician's Colleagues/Local Medical Community Towards Consultant. Norms that the physician's colleagues and local medical community have established for using that consultant for the patient.

Physician's Satisfaction with Previous Referrals to Consultant. Satisfaction with results; the way patients were treated; the way the physician was treated.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Use of Illness-Related Ambulatory Physician Services in Finland

PERTTI KEKKI, M.D., SC.D., DCM(EDIN)

The total utilization rate of ambulatory illness-related physician services in Finland is higher in big urban health center areas than in rural areas. When only the health center physician utilization rates were studied the finding was opposite. A notable finding related to the impact of the national health planning were the surprisingly equal average utilization rates of public sector health services (health center physician and general hospital outpatient services combined) in both areas. In fact, use of cheaper public services was greater in rural than in urban areas. The difference in total use rates was almost entirely due to the larger use of more expensive private physician services in urban areas. At present, there is a great pressure for increasing the supply of health center services in urban areas to the same level as in rural districts. One of the fundamental questions seems to be the pricing of the two service sectors—public and private.

IN ALMOST ALL of the industrialized countries, efforts have been made to improve the availability and accessibility of health services in order to be able to provide the necessary services to the population, regardless of the place of residence or income status.

The common finding, however, has been that even in countries where the above goals have been stated in the legislation, there are continous inequalities. In England, for example, it was found that the inequalities between the regions had increased rather than decreased during the first 21 years of the N.H.S. Those regions which previously had many services available, had gained more, or the differences had remained the same.[1] Hart[2] presented the Inverse Care Law, according to which, in England, the availability of the high quality physician services was inversely related to the need of the population for those services.

In Finland, the goals of the universal sickness insurance (in effect from 1964)[3] and the Community Health Care Act of 1972, which emphasized the importance of primary care and created the health center network covering the whole country,[4,5] were to make the population groups living in various parts of the country equal as for the possibilities to use health services.

The development of health services after 1972 is based on five-year health planning, according to which the public sector health resources are being allocated. In this planning mechanism, there is an in-built evaluation system. However, despite the various planning activities, the factual distribution of the services or their use have not been widely examined. This is due to the gap in the planning system. To get a real picture

---

Institute of Public Health Science, University of Tamperg and Department of Primary Medical Care, University of Helsinki.

Supported in part by grants from the Finnish government and the Yrjö Jahnsson Foundation (Finland).

The basic work for this study was carried out in collaboration with the Research Institute for Social Security of the Social Insurance Institution of Finland at the Division of Health Care Organization of the Johns Hopkins University School of Hygiene and Public Health in Baltimore, Maryland.

Reprint requests: Pertti Kekki, M.D., Sc.D., DCM(Edin), Assistant Professor (Health Services Administration) University of Tamperg and Acting Professor of Primary Medical Care, University of Helsinki. Department of Primary Medical Care, University of Helsinki, 2. Linja 4C, SF-00530 Helsinki 53, Finland.

0025-7079/82/0800/0797/$01.40 © J. B. Lippincott Co.

of the developments and utilization of the various services at the primary level is impossible, since the units examined are large aggregates (provinces or central hospital districts). At the provincial level there are, of course, currently better possibilities for studying smaller units, such as the individual health center areas. The difficulty there lies in the data collected routinely because these do not allow for in-depth analyses as some more specific data are not routinely collected.

In some studies (e.g., Tuomikoski[6], it has been suggested that planning has not brought about the solution in the distribution or clustering of the services.

Apart form the public sector services (health centers and the great majority of hospitals), people in Finland can also use the services of the private sector. The basic difference between the health center and private sector illness-related physician services is that the physicians working in the former are salaried medical officers and the physicians working in the latter are usually hospital physicians who when working in the hospitals are full-time salaried staff but, who after their regular working hours, turn to private practitioners working on a fee-for-service basis in private medical centers or group practices outside the hospitals. The majority of the private practioners are specialists whereas the health center physicians are usually general practitioners. In 1975, about 60% of all physicians working in hospitals were in private practice, the proportion being probably greater for physicians working in general hospitals. Only about 10% of all physicians in the country were full-time private practitioners.

Sickness Insurance reimbursed in 1975 about 44% of the private physician fee to the patient afterwards, the average fee being about 31 FIM(US$ 7.50). In health centers, the patient paid about US$0.80 for a health center physician visit in 1975. The differences will be even greater if the fees for diagnostic services (lab, x-ray) are taken into account. From the beginning of 1976, laboratory and x-ray services in health centers have been free to the patient, and since the beginning of 1981 the physician visit also is "free."

Goals

The general goal of the present study is to increase the knowledge and understanding of health services in Finland. For planning purposes, it is considered useful to see whether a mechanism of national health planning, acting in a well-defined health services system in a developed country, has been able to increase equity with regard to use of physician services in ambulatory care.

The specific aims of this research were: 1) to study whether, in 1975, there still were considerable differences between urban and rural areas with regard to the volume of use of ambulatory physician services; 2) in case differences that exist, to study also the variation of use within the groups of urban and rural areas and the possible reasons for these differences; and 3) to correlate the findings with the goals of the national health policy and health planning.

It is especially important to note that, in this study, we use as the units the basic units of the Finnish health services—the health center areas, in contrast to other studies which have used the basic administrative units—the communes or local authority areas. Most of the health center areas are composed of more than one commune. This fact makes the findings more reliable from the health planning point of view.

Materials and Methods

The material for this study comes from a data set which has been described in detail elsewhere.[7] The material was processed so that the final data set included all of the basic units of the Finnish health services, the health center areas existing in 1975 (216 in total). From each health center area, a large number of resource, demographic, social structure, utilization, and other variables were included. The data, thus in each case, referred to the mid-year population of the particular health center area. The original data were precollected and came from various sources such as the Social Insurance Institution, the Finnish Hospital Association, the Central Statistical Office, the Office of the Cabinet, and the Finnish Medical Association.

In 1975, there were ten health center areas in Finland with a population of 60,000 or more. Conversely, 72 per cent of all the health center areas had a mid-year population of 20,000 or less, and 85 per cent of all 30,000 of less. This study includes 38 health center areas, of which 19 represent the big population centers. These include eight of the areas with a population of 60,000 or more. All of these big centers except one (Vantaa) have a central hospital located in the area (five of them have a university teaching hospital). The other 19 centers are mainly rural health center areas with a population 20,000 or less. This paper examines the use of ambulatory illness-related physician services by the populations of these health center areas. Figure 1 shows the distribution of the centers of the areas included in this study.

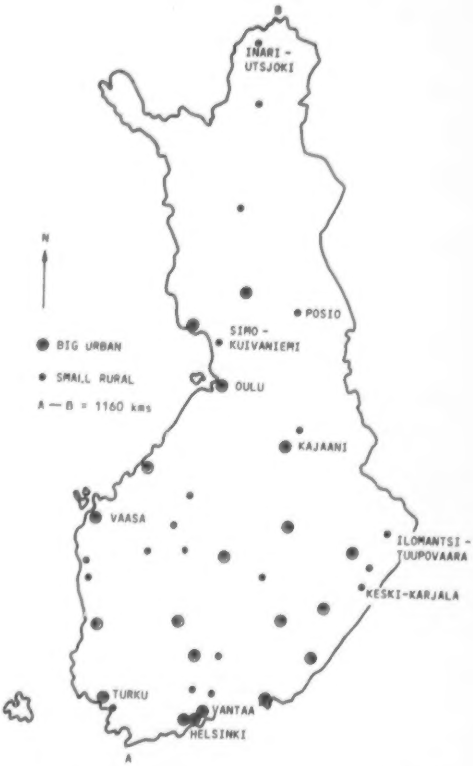


FIG. 1. Map of Finland showing the distribution of the centers of the health center areas included in the study.

Results

Table 1[†] shows the coverage of illness-related physician services for the health center physicians and private physician services. This is done by presenting the proportions of the populations in each health center area that did not use these services in 1975 (the proportion of the nonusers).

The differences between the health center areas were considerable. Whereas 31 per cent of the population in Helsinki (the capital city) used health center physician services in 1975, 60 per cent of the population of the rural health center area of Keski-Karjala (in Eastern Finland) used these services. Conversely, about 43 per cent of the population in Helsinki used private physician services, whereas in the health center area of Keski-Karjala, only 14 per cent of the population received these services. It is, however, interesting to notice that the total percentages of users in both places are equal (74 per cent). In

---

[†] In these tables the coefficient of variation (C%) also has been calculated. $C\% = 100 \times SD/X$. This coefficient is measure of relative variability whereas the standard deviation is a measure of absolute variability.

TABLE 1. Percentages of the Nonusers of Services in the Population by Health Center Area

Large Health Centers	HC-Physician	Private Physician	Small Health Centers	HC-Physician	Private Physician
Helsinki	68.8	57.3	Alajärvi	51.8	72.0
Oulu	74.5	59.9	Jalasjärvi	41.2	88.7
Turku	83.3	49.5	Närpiö	49.6	87.0
Tampere	75.2	58.6	Perhonjokilaakso	51.9	83.3
Kuopio	66.6	55.9	Länsi-Turunmaa	54.6	68.2
Jyväskylä	66.3	62.1	Kiihtelys-Pyhäselkä	48.6	87.3
Joensuu	55.4	74.4	Keski-Kainuu	48.0	84.2
Pori	72.6	53.1	Ilomantsi-Tuupovaara	44.8	92.5
Savonlinnanseutu	53.6	74.6	Keski-Karjala	39.3	85.7
Mikkelinseutu	56.1	72.2	Simo-Kuivaniemi	65.6	80.4
Kotka	61.9	62.2	Sodankylä	39.4	91.4
Vaasa	77.4	65.0	Posio	40.9	87.8
Rovaniemi	58.1	69.4	Ähtärinjärvi	59.0	73.7
Lappeenranta	57.4	69.6	Kristiinanseutu	71.5	70.1
Hämeenlinnanseutu	63.3	62.6	Tuusula	46.0	75.2
Kajaani	54.5	72.1	Inari-Utsjoki	42.0	90.7
Kokkolanseutu	66.2	66.6	Nurmijärvi	44.0	73.8
Vantaa	59.8	58.3	Lamminseutu	44.3	83.3
Kemi	58.5	74.9	Kangasniemi	44.0	84.4
$\overline{X}$	64.7	64.1	$\overline{X}$	48.8	82.1
SD	8.7	7.7	SD	8.8	7.6
C (%)*	13.4	12.0	C (%)	18.0	9.3

* C (%) = coefficient of variation. $C = 100\left(\frac{SD}{\overline{X}}\right)$.

Ilomantsi-Tuupovaara health center area, another rural area situated in North Karelia, only 7.5 per cent of the population had contacted a private physician, and in Lapland (in the Northern part of the country) nine per cent of the population of the Inari-Utsjoki rural health center area contacted a private physician in 1975. When the mathematic means of the utilization rates for the health center and private physicians are calculated, they show that in the small health center group roughly about 50 per cent of the population used health center physician services, compared with the average of about 35 per cent in the group of big centers. As for private physicians, roughly 19 per cent of the populations of the small centers used these services, compared with the about 36 per cent in the big centers. These figures compared well with the averages for the entire country: in 1975, about 43 per cent of the total population used health center physician services—about 37 per cent of the population living in urban areas and 51 per cent of the population living in rural areas. About 30 per cent of the total population used private physician services, about 37 per cent of the population in urban areas and 20 per cent in rural areas. About 41% of the total population did not use either of these two services in 1975. This indicates that some overlapping takes place. Some people have used both the HC and private physician services. In the entire country, about 29 per cent of the population used only health center physician services in 1975, 16 per cent used only private physician services and 14 per cent used both. Over half of the visits made to private physicians were made to specialists. Table 2 presents the utilization rates by health center area for health center physician visits, private physician visits, and outpatient

TABLE 4. General Surgeon (GS) and General Practitioner (GP) Rankings of the Decision Where to Refer Factors (15—Most Important, 1—Least Important)

	General Surgeons												General Practitioners														
Factors	1	2	3	4	5	6	7	8	9	10	11	Average	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Average
Quality of patient management	15	15	15	11	15	3	15	15	14	15	15	15	15	14	15	14	14	12	14	15	15	15	15	13	15	14	15
Patient results	14	14	13	4	14	7	14	13	13	14	14	14	14	15	13	15	15	4	15	11	14	10	8	14	13	12	13.5
Individualized patient management and care	13	13	6	10	13	5	11	8	15	13	13	12	13	12	14	13	13	11	13	14	13	14	14	8	8	13	13.5
Convenience to the patient	4	7	5	13	5	8	9	4	9	9	11	7	6	4	8	2	5	8	4	12	5	4	13	9	3	8	6
Cost to the patient	3	4	2	3	4	2	3	3	8	6	10	3	5	3	5	3	7	9	3	10	1	3	11	10	11	5	5
Degree patient's preferences are met	9	6	3	12	12	15	12	7	10	10	8	10	9	10	9	12	6	5	7	13	12	5	12	11	9	15	11
Physician's loss of patient/family/income	1	1	1	1	1	4	1	2	1	5	1	1	3	1	2	5	1	1	1	1	2	1	3	1	1	1	1
Reciprocations physician receives	6	2	10	2	3	6	5	6	2	4	3	4	2	7	3	4	2	7	6	2	4	6	9	5	7	2	3
Communication physician receives	5	5	7	5	9	10	8	9	6	3	9	6	7	5	7	10	4	10	5	8	11	9	10	4	10	9	7
Respect/courtesy physician receives	10	10	4	6	2	9	7	5	7	2	2	5	12	6	1	8	3	6	9	4	10	7	2	3	4	3	4
Previous use of and satisfaction with consultant by patient/family	7	8	11	14	6	14	4	14	3	7	12	9	4	8	10	7	9	14	10	6	9	11	4	12	6	6	9
Attitudes of patients toward consultant	11	9	9	15	7	13	6	10	4	8	5	8	8	11	11	6	10	2	8	7	6	8	5	15	5	10	8

visits to general hospitals, as well as the total ambulatory physician utilization rates.

The health center physician visit rate varies from 361/1000 to 2122/1000 in Posio (a rural health center area in the province of Lapland). Knowing also the population bases of the areas the findings suggest that the smaller the central hospital city the higher the health center physician visit rate seems to be (the smaller the private sector and probably the more developed the health center sector). The weighted average for the health center physician visit rates in the areas with more than 60,000 inhabitants is 643/1000, and for the first ten of the smaller health centers in the list this figure is 1440/1000. The weighted average for the private physician visit rates in the large center group is 1389/1000, compared with the 344/1000 in the small center group. In the big center group, the average visit rate to the outpatient departments of general hospitals was 929/1000 while this rate for the small centers is 329/1000. When all these rates are added up for the 19 large and small centers, the resulting total population weighted average use rates are 2819/1000 in the large centers and 1931/1000 in small centers. The difference is almost one visit per person in favor of the big population centers. The emphasis in the big centers is strongly on the private physician sector and, on the other hand, on use of the outpatient services of the big hospitals in these areas. In these areas, the population, as the consequence, uses considerably more specialist services than in the small health center areas (probably for primary care?). The highest total use rate among the health center areas listed here was in Turku (the third largest city in the country on the Southwestern coast) 3.3 visits/person. The next highest was in Helsinki, 3.2 visits/person. The lowest the figure was in the rural health center area of Simo-Kuivaniemi in the sparsely populated Lapland, 1.7 visits/person.

Table 3 presents the utilization rates per 1000 people in three age groups for health center and private physician services. (Hospital outpatient visit rates were not available by age groups.) This table shows interesting differences between the two groups, but also within the groups. For example, in the large centers, the health center physician visit rate for children varies from 855 in Vantaa to 360 in Oulu (a university city in the North). In the small health center group, the range is from 1227 in the area of Inari-Utsjoki to 355 in Simo-Kuivaniemi, both of these areas in Lapland. The rate of private physician contacts varies in the large center group from 829 in Turku to 116 in Kajaani (a relatively small central hospital city in the North) and in the small health center group from 302 to 23. Similar differences can be found in the other age groups.

When the total age-specific utilization rates (health center and private physician use combined) are calculated the variation seems to be smaller in the group of large centers. The highest combined average contact rates are found in the group of the eight large centers having a population of 60,000 or more. In this group, the city of Vantaa has the highest health center physician visit rate for children of all the 38 health centers under study. (Vantaa is a big neighboring city of Helsinki.) Also, in other age groups, the health center visit rates in Vantaa rank high. In the age group of 16–64 years, Vantaa is third after Posio, in Lapland, and Nurmijärvi, near Helsinki. In the age group for the elderly Vantaa is second only to Posio (Posio 3.9 visits/person versus 3.6 visits/person in Vanaa).

In the central hospital cities, the smallest use rates are found in the city of Vaasa (on the West coast about 400 km from Helsinki), where the total utilization rates in each age group are the lowest in the group. The reason for this seems to be the low health center physician use rates, which the private sector is not able to compensate. Of the small health centers, the worst situation seems to be in Simo-Kuivaniemi, where the utilization figures for each age group are lowest.

The most interesting health center of those studied seems to be the health center area of Posio, where the utilization rates for the working age population and the elderly are the highest among the 38 centers, and the utilization rate for the children is also at a relatively high level. The high use rates in Posio are mainly due to the high health center physician visit rates (Table 3). Since in Posio especially the rates for the working age population and the elderly are in their own class, the next step in this study was to examine more closely the health center physician visits in order to find perhaps some clues about the possible reasons for the big differences.

A high use rate may be a consequence of a high coverage of services i.e., the high proportion of the population who has used the services or it may also be a consequence of a high visiting rate of the same persons (frequent visitors). According to Table 1, the coverage of the health center physician services in Posio is not the largest. It is among the largest, but there are areas where the coverage is higher or at the same level as in Posio. Because of this finding, the distribution of such variables as the number of visits per visitor to health center physicians and the proportion of visitors among all those who visited who made seven or more visits to health center physicians in 1975 were studied.

Table 4 shows, among other things, that when in the city of Vaasa there were an average of 1.7 visits/visitor to health center physicians; this figure in Posio was 3.6. When the proportion of the high utilizers was 1.0 per cent in Vaasa, it was 13.2 per cent in Posio. Table 4 also suggests that in all those health center areas where the average health center physician visit rate/

TABLE 4. Specific Characteristics of Health Center Physician Visits by Health Centers

Large Health Centers	Visits per Visitor	Proportion Visitors with Seven or More Visits	Small Health Centers	Visits per Visitor	Proportion Visitors with Seven or More Visits
Helsinki	2.3	4.3	Alajärvi	2.7	6.6
Oulu	1.9	2.3	Jalasjärvi	3.0	9.0
Turku	2.1	3.7	Närpiö	3.0	9.4
Tampere	1.9	2.3	Perhonjokilaakso	2.4	4.4
Kuopio	2.2	3.7	Länsi-Turunmaa	2.5	5.5
Jyväskylä	2.2	3.5	Kiihtelys-Pyhäselkä	2.3	3.5
Joensuu	2.3	3.6	Keski-Kainuu	2.7	6.5
Pori	2.1	2.8	Ilomantsi-Tuupovaara	2.7	6.8
Savonlinnanseutu	2.3	3.9	Keski-Karjala	3.3	11.1
Mikkelinseutu	2.5	5.3	Simo-Kuivaniemi	2.1	2.4
Kotka	2.1	2.7	Sodankylä	2.9	8.3
Vaasa	1.7	1.0	Posio	3.6	13.2
Rovaniemi	2.2	3.0	Ähtäarinjärvi	2.2	3.0
Lappeenranta	2.3	4.0	Kristiinanseutu	2.6	6.3
Hämeenlinnanseutu	2.3	4.0	Tuusula	2.8	7.3
Kajaani	2.7	5.9	Inari-Utsjoki	2.9	8.6
Kokkolanseutu	2.2	3.6	Nurmijärvi	2.8	6.8
Vantaa	2.3	4.3	Lamminseutu	2.6	5.5
Kemi	2.7	8.2	Kangasniemi	3.2	10.6
$\overline{X}$	2.2		$\overline{X}$	2.8	
SD	0.2		SD	0.4	
C (%)	9.1		C (%)	14.3	

visitor is 3.0 or more, the proportion of those visitors who made seven or more visits is about 10 per cent or more. The finding also seems to correlate well with the visit rates in Table 2. This finding seems to suggest that the high utilization rates result from the fact that the same persons visit frequently. The differences between the big and small health center areas are clear but the differences are also considerable within the group of the small health centers.

In order to see whether this explanation is valid we studied a group of 16 health centers, eight of which represent the lowest average health center physician use rates per person in 1975 in the country, and the other eight the highest utilization rates.

In Table 5, the health center physician population ratios are also included. From the figures in the table it can be concluded that there seems to be a relatively good correlation between the number of visits per visitor and the proportion of high utilizers among these visitors. (There also seems to be a good correlation between these two variables and the supply of health center physicians.

The findings thus indicated that:

1) there are differences in utilization patterns of ambulatory physician services between urban and rural areas. When only the public ambulatory physician service use rates (HC physician and outpatient department services which both are controlled by the national planning) are studied the use per population seems to be higher in rural than in urban areas ($p <$ 0.02, t-test), whereas looking at the total use of illness-related ambulatory physician services (private sector use added), the populations of the urban areas seemed to use more services ($p < 0.001$, t-test).

2) When variations within the groups were studied it was shown that there were relatively large variations in both groups and that for some reason the variability seemed to be greater within the group of the rural health centers.

3) In light of the findings it can be stated that the national health planning mechanism which was put into effect in 1972 seems to have succeeded in leveling off the great differences in utilization of health services existing before 1972 between the urban and rural areas. This was one of the main goals of the 1972 legislation and the reorganization of the health services in Finland based on it.

TABLE 5. Health Center Physician Visits per Visitor in Health Centers Representing the Lowest and the Highest Rates in Finland 1975

Centers with Lowest Rates				Centers with Highest Rates			
Health Center	Visits per Visitor		High Utilizers (%)	Health Center	Visits per Visitor		High Utilizers (%)
Vaasa	1.7	(1.0)	1.0	Ikaalinen	3.9	(3.6)	17.4
Pietarsaari	1.8	(1.3)	1.2	Janakkala	3.7	(2.6)	15.0
Seinäjoki	1.8	(1.4)	1.2	Hämeenkyrö	3.7	(3.5)	14.4
Mustasaari	1.9	(1.2)	1.4	Posio	3.6	(4.1)	13.2
Lahti	1.9	(1.8)	1.6	Ruovesi	3.5	(4.4)	13.4
Oulu	1.9	(2.1)	2.3	Hanko	3.5	(3.9)	12.9
Tampere	1.9	(1.6)	2.3	Parkano	3.4	(3.5)	12.8
Kauniainen	1.9	(1.5)	2.7	Harjavalta	3.4	(2.9)	12.2
$\overline{X}$	1.9	(1.6)	1.7	$\overline{X}$	3.6	(3.6)	13.9

Discussion

According to the findings presented there were considerable differences in use of physician services between the different types of health center areas in Finland in 1975 by type of services used. In large population centers, considerably more services were used per person. There seemed to be an especially great difference in use of private physician services—use of specialist services—despite the universal sickness insurance. In this respect, people living in the rural areas seemed to be in an inferior position. It is not certain on the basis of the data available, however, whether people in the big population centers use specialists for true specialist medical service needs or whether these specialists are used for first-line primary care due to the lack of ordinary primary care physicians (Fig. 2). In case the demand in cities reflects a real need for specialist care (which is not probable) then one of the priorities should obviously be the increase of the availability of specialist services also to the populations of the rural areas. An effective method for doing this in Finland would be the development of the availability and accessibility of the outpatient services of general hospitals in the developing regions. This is because it is impossible to force private physicians to these regions but it is possible to allocate positions to general hospitals located in these regions, and it is also possible to designate the positions for the outpatient departments. This would improve the possibilities of consultation services for health centers. Taking into account the development in the number of physicians in the country as well as the fact that the great majority of the physicians are salaried staff of public institutions, it is almost certain that in a relatively near future even specialists will go where the positions are. It is understandable from the figures that looking only at the utilization rates for health center physicians one may get an

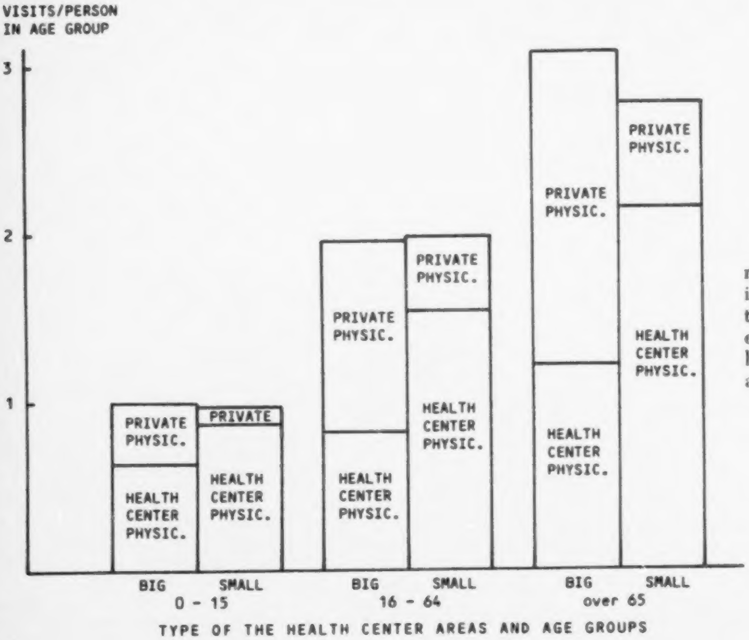


FIG. 2. Use of illness-related physician services in ambulatory care (hospital outpatient depts excluded) by type of the health center area and by age groups in 1975.

impression that the populations of the large urban health centers are in a worse position from the equity point of view. Worse here referring to the availability of cheap (to the patient) health center physician services. As mentioned earlier, the original out-of-pocket cost of a physician visit in 1975 was about 7.5 times less expensive in the health center than in the private sector. In case lab tests or x-ray examinations were ordered the difference was much greater.

Thus, the basic problem in the big centers seems not to be the lack of services but 1) the skewness of the structure of the supply which, perhaps, results in use of specialist services when the demand would be satisfied by general practitioners if enough were made available; 2) the great difference of the out-of-pocket costs of the services in the two sectors—public and private. It is likely that when the choice is between "free" health center physician services and private physician services, for which the patient has to pay considerably more, the demand is for health center services and the population considers its position worse in the big centers because the low-cost services are not as readily available as in the rural areas.

In the small health centers the structure of the supply is skewed toward another direction: there are basic services available but there may be barriers for use of specialist services. When the average visit rates per visitor were calculated, differences were also found within the groups. Especially notable, these differences were among the small health centers. The great differences probably are partly due to differences in treatment habits and partly also due to differences in available resources between the various health centers. Where the visiting rate was high, without exception there was also a large proportion of high utilizers to be found. This was confirmed by examining the 16 health centers with the highest and lowest health center physician utilization rates per visitor in the country. The observed high utilization rates per visitor and the large proportions of the high utilizers connected with them can at least partly be explained by the great number of repeat visits (whether needed or not) in these health centers. The basic reasons for these differences cannot, however, be explained by the data available, since more information would be needed on the consulting practice, the patient mix, organization of resources, etc. from these health centers. These findings, however, stress the importance of examining the effectiveness and efficiency aspects of the health services. On the basis of the figures presented, it cannot be stated what can be considered an excessively high visiting rate or what the rate should be. It would be, however, interesting for further analyses to be able to distinguish between the first and repeat visits, which is, at the moment, impossible. Also it would be interesting to study why the variability in use (and probably also in supply) of health center physician services is so great even within the groups.

Finally, a notable finding pointing to a favorable impact of national health planning can be seen in Figure 3, which shows a remarkable equality between the two groups of health center areas, as for the public services which are controlled by planning. The general impression from the findings is that by 1975 the coverage of the basic physician services in the rural areas in Finland was relatively good. However, although the health center services in big cities in general seemed to be undeveloped, there is no question about the clustering of services in them, which also shows in the bigger total utilization rates. However, based on the strong demand for the development of the health centers in the bigger urban areas existing at present, it seems that we are now approaching a point where a serious consideration should take place as to how to pro-

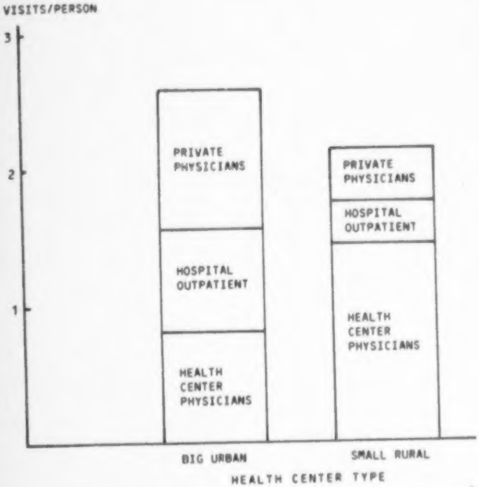


FIG. 3. Use of illness-related physician services in ambulatory care by type of the health center area in 1975.

ceed. The underlying dilemma is how to avoid increasing the inequality in resource supply and availability in favor of the urban areas again. One of the basic problems being linked with the difference in pricing of the health center and private sector services. The characteristics of the two service sectors are also different (general practitioners who are salaried versus specialists working on a fee-for-service basis who, at the same time, are salaried full-time hospital physicians). So, we are not comparing like with like. At this point it is impossible to tell what the development will be. The decisive factor will be the taste of the consumer.

Acknowledgment

The author wishes to express his thanks to Dr. Robert L. Parker of the Department of International Health at Johns Hopkins School of Hygiene and Public Health for his comments on the original manuscript.

References

1. Griffiths DAT. Inequalities and management in the N.H.S. Hospital 1971;229.

2. Hart TJ. The inverse care law. Lancet 1971;i:405.

3. Purola T, Kalimo E, Nyman K, Sievers K. National health insurance in Finland: its impact and evaluation. Int J Health Serv 1973;3:69.

4. Kekki P. Same goals-different philosophy. Br Med J 1976;1:204.

5. Kekki P. General practice in Finland. J R Coll Gen Pract 1976;26:853.

6. Tuomikoski H. Terveyspalvelusten tarjonta ja tavoitettavuus Suomessa 1964–1976. (The Supply and Availability of Health Services in Finland 1964-76.) Thesis. Dept. of Social Policy, University of Helsinki, November 1978.

7. Kekki P. Analysis of relationships between availability of resources and use of health services in Finland. A cross-sectional study. Med Care 1980;18:1228.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Visits to Physicians by Asian/Pacific Americans

ELENA S. H. YU, PH.D.,* AND BEULAH K. CYPRESS, PH.D.†

The purpose of this paper is to disseminate some of the findings from the 1979 National Ambulatory Medical Care Survey (NAMCS) where, for the first time in its history, data are available on the characteristics of visits to office-based physicians by Asian/Pacific Americans. The NAMCS Program is the only source for such data which is based on a national probability sample survey of physician–patient encounters. Among the findings are: 1) no significant difference was observed between Asian/Pacific Islanders and the white majority with regard to the patient's sex, return-visit ratio, or the kinds of diagnostic and therapeutic services rendered by the sampled physicians. 2) Children appear to be the major utilizers of ambulatory medical care among Asian/Pacific Americans. 3) In every age group, the visit rate to office-based physicians is lower for Asian/Pacific Americans when compared with white Americans. 4) A significantly smaller percentage of Asians or Pacific Islanders, in contrast to other specified race/ethnic groups, had visited a physician for injury or poisoning, and a substantial proportion of their visits were apparently made for preventive care. 5) Significantly fewer visits were made by Asian/Pacific Americans to the office of a surgeon or a psychiatrist. The demographic antecedents and cultural underpinnings of these findings are discussed.

BETWEEN 1970 AND 1980, the total number of Asians and Pacific Islanders in the United States more than doubled, from 1.5 to 3.5 million.[1] This dramatic increase came about largely through legal immigration made possible by the 1965 Amendments in the U.S. immigration law (which took effect in 1968) and the passage of the Refugee Act following the fall of South Vietnam and other parts of IndoChina. The surge in Asian/Pacific Islander population during the 1970s has exceeded that found for all other ethnic groups including persons of Spanish origin. If the present trend continues, even greater increase in the Asian/Pacific American population is expected by the 1990s. Yet, little is known about the health status, services utilization, and ambulatory medical care of this growing minority.

In 1979, however, the National Ambulatory Medical Care Survey (NAMCS) conducted by the National Center for Health Statistics (NCHS) allowed for the identification of "Asian/Pacific Islanders" on the questionnaire filled out by a probability sample of physicians selected for the survey in the coterminous United States. Thus, for the first time in the history of the NAMCS Program, data are available for minorities other than black Americans re-

---

* Division of Analysis, National Center for Health Statistics, Department of Health and Human Services, Hyattsville, Maryland.

† Division of Health Care Statistics, National Center for Health Statistics, Department of Health and Human Services, Hyattsville, Maryland.

Address for reprints: Elena S. H. Yu, Ph.D., Division of Analysis, National Center for Health Statistics, FCB#3, Room 2-27, 3700 East West Highway, Hyattsville, Maryland 20782.

0025-7079/82/0800/0809/$01.40 © J. B. Lippincott Co.

garding the characteristics of their visits to office-based physicians. Although there are methodologic problems in the interpretation of "Asian/Pacific Americans" because the label actually encompasses a heterogeneous population from diverse cultures,‡ the importance of obtaining at least some preliminary baseline information on this ethnic group seems to override the disadvantages. To date, there are no systematic studies on the ambulatory medical care of Asian and Pacific Americans. The NAMCS data appear to be the only source of information for such type of care, and it is the only data set which is based on a national probability sample survey of physician–patient encounters.

The purpose of this paper is to disseminate some of the findings from the NAMCS Program where the number of Asian/Pacific Americans for selected characteristics are sufficiently large as to permit meaningful tabulations. Data for black and white Americans are also included in the presentation that follows for the purpose of comparison. It is hoped that the general availability of these data, however limited, will stimulate greater interest among investigators in the field of health services to conduct more systematic research on the health care of minorities.

Sample Design of NAMCS

The target population of NAMCS encompasses office visits within the coterminous United States made by ambulatory patients to nonfederally employed physicians who are principally engaged in office practice, but not in the specialities of anesthesiology, pathology, or radiology. Also excluded were physicians principally engaged in teaching, research, or administration. Likewise, telephone contacts and nonoffice visits were not included in the scope of the survey. A multistage probability sampling design is used which involves probability samples of primary sampling units (PSUs), physician practices within PSUs, and patient visits within practices. For 1979, a sample of 3,023 non-Federal, office-based physicians was selected from master files maintained by the American Medical Association and American Osteopathic Association. The physician response rate was 71.8 per cent. Sampled physicians were asked to complete a questionnaire called Patient Records for a systematic random sample of office visits taking place within their practice during a randomly assigned weekly reporting period. A total of 45,351 Patient Records representing an estimated 556,313,431 office visits were completed by sampled physicians.[2] Of these, 442 records, representing an estimated 5,559,524 visits, characterized patients as "Asian/Pacific Islanders."

Limitations of the NAMCS Data

Statistics from the 1979 National Ambulatory Medical Care Survey were derived by a multistage estimation procedure, which produces essentially unbiased national estimates[2] and has three basic components: 1) inflation by reciprocals of the probabilities of selection, 2) adjustment for nonresponse, and 3) a ratio adjustment to fixed totals. Notwithstanding these safeguards, caution is warranted in the interpretation of the NAMCS data for Asian/Pacific Americans. First, ethnic identification is not made by the patients themselves but by the sampled physicians or their assistants, based on their prior knowledge of patients or observations. Second, differences exist between Asian/Pacific American subgroups and within each national origin grouping with regard to their time of first arrival in the United States, reasons for immigration, socioeconomic status, and educational attainment, to mention only a few factors.[3,4] However, insofar as Asian/Pacific Americans tend to be viewed as "alike" and are treated accordingly in the health care setting, it is legitimate for some data analysis purposes to examine their health utilization behavior as a group. Third, due to the nature of the sampling design and the resultant disparity in sample sizes between specified race/ethnic groups, the estimates of the volume and characteristics of visits to office-based physicians are more precise for the majority white population than those obtained for Asian/Pacific Americans. Despite this caution, the NAMCS data for Asians and Pacific Islanders are sufficiently precise as to be useful in two important senses: 1) by way of contrast with similar data obtained from the white majority, and 2) by providing a preliminary estimate of the characteristics of the minorities sampled in the NAMCS Program.

---

‡ Broadly speaking, the term Asian/Pacific Americans designates residents of the United States from the following countries and territories: *East Asia*: China (including Taiwan and Hong Kong); Japan and Korea. *Southeast Asia*: Philippines, Vietnam, Cambodia, Laos, Thailand, Malaysia, Singapore, and Indonesia. *Indian Subcontinent or South Asia*: India, Pakistan, Bangladesh, Sri Lanka, and Burma. *Pacific Islands*: Hawaii, Guam, Samoa, Tonga, Fiji, and other Micronesian Islands. In practice, however, only a few major groups originating from the Asian-Pacific triangle are identified in federal statistics. These major groups include persons identified as Chinese, Japanese, Filipinos, Koreans, IndoChinese (Vietnamese, Cambodians, and Laotians), Hawaiians, Guamanians, and Samoans. Some Census Bureau reports have also identified (East) Indians separately.

NAMCS Data for Asian/Pacific Americans

In 1979, the majority of the physician visits made by Asian/Pacific Americans occurred in metropolitan areas (93.8 per cent) rather than in nonmetropolitan areas (6.2 per cent). In contrast, less than three-quarters of white American patient visits occurred in the metropolitan areas (Table 1). This pattern of concentration found for Asians and Pacific Islanders is consistent with their geographic distribution in the United States. For instance, an independent estimate based on the National Health Interview Survey (NHIS) conducted in 1978 by the National Center for Health Statistics indicates that nearly 92 per cent of Asian/Pacific Americans reside in standard metropolitan statistical areas.

In the NAMCS data, no discernible difference was observed between Asian/Pacific Americans and the white majority with regard to the patient's sex. In both groups, proportionately more females (60 per cent) than males (40 per cent) visited physicians during the survey period. In general, 45 per cent of the visits by Asian/Pacific Islanders and 38 per cent of those by white Americans were for new problems. The ratio of return visits to new

TABLE 1. Per Cent Office-based Physician Visits in Metropolitan and Nonmetropolitan Areas for Specified Race/Ethnic Group: United States, 1979

	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent	100.0	100.0	100.0	100.0
Metropolitan area†	73.4	72.5	81.8	93.8
Nonmetropolitan area	26.6	27.5	18.3	6.2

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the Coterminous United States).

† Standard metropolitan statistical area as defined by the U.S. Bureau of the Census.

TABLE 2. Illness-related Visiting Rates/1000 in Various Ambulatory Care Sectors by Health Center Area

Large Health Centers	HC-Physician Visits	Private Physician Visits	Outpatient Visits	Total	Small Health Centers	HC-Physician Visits	Private Physician Visits	Outpatient Visits	Total
Helsinki	739	1458	1046	3243	Alajärvi	1321	668	297	2286
Oulu	493	1153	1069	2715	Jalasjärvi	1773	179	237	2189
Turku	361	1648	1291	3300	Närpiö	1545	210	284	1914
Tampere	477	1235	767	2479	Perhonjokilaakso	1177	320	382	1879
Kuopio	738	1305	674	2717	Länsi-Turunmaa	1142	816	312	2270
Jyväskylä	756	1026	533	2315	Kiihtelys-Pyhäselkä	1185	245	665	2095
Joensuu	1021	603	801	2425	Keski-Kainuu	1391	291	362	2044
Pori	564	1555	628	2747	Ilomantsi-Tuupovaara	1506	118	384	2008
Savonlinnanseutu	1062	589	531	2182	Keski-Karjala	1988	248	118	2354
Mikkelinseutu	1107	638	570	2360	Simo-Kuivaniemi	715	381	579	1675
Kotka	812	1070	700	2582	Sodankylä	1746	122	110	1978
Vaasa	378	1063	839	2280	Posio	2122	160	132	2414
Rovaniemi	910	776	612	2298	Ähtärinjärvi	887	674	342	1903
Lappeenranta	996	779	746	2518	Kristiinanseutu	733	773	471	1977
Hämeenlinnanseutu	862	1084	572	2518	Tuusula	1490	629	471	2590
Kajaani	1206	688	714	2608	Inari-Utsjoki	1703	122	126	1951
Kokkolanseutu	742	876	890	2508	Nurmijärvi	1550	654	415	2619
Vantaa	925	1257	613	2795	Lamminseutu	1439	342	236	2017
Kemi	1145	564	783	2492	Kangasniemi	1780	271	220	2271
$\overline{X}$	805	1019	757	2583	$\overline{X}$	1431	380	323	2128
SD	259	337	201	295	SD	154	57	24	251
C (%)	32.2	33.1	26.6	11.4	C (%)	10.8	15.0	7.4	11.8

TABLE 3. Health Center and Private Physician Visit Rates per 1000 by Age Groups in 1975

Large Health Centers	HC-Physician Visit			Private Physician Visit			Small Health Centers	HC-Physician Visit			Private Physician Visit		
	0–15	16–64	65–	0–15	16–64	65–		0–15	16–64	65–	0–15	16–64	65–
Helsinki	607	676	1275	631	1570	2005	Alajärvi	765	1445	1911	124	773	1349
Oulu	360	497	948	472	1284	2368	Jalasjärvi	948	1972	2246	40	235	154
Turku	440	256	842	829	1722	2629	Närpiö	861	1556	2345	81	236	264
Tampere	487	452	595	430	1331	2171	Perhonjokilaakso	572	1243	2336	61	412	399
Kuopio	581	765	945	414	1412	2822	Länsi-Turunmaa	1028	1085	1515	302	862	1284
Jyväskylä	645	744	1151	415	1125	1839	Kiihtelys-Pyhäselkä	573	1259	2091	58	296	359
Joensuu	624	1079	1485	150	673	1286	Keski-Kainuu	709	1542	2165	63	371	350
Pori	503	560	732	512	1814	2281	Ilomantsi-Tuupovaara	872	1601	2184	34	144	139
Savonlinnanseutu	723	1112	1508	129	657	1198	Keski-Karjala	1072	2105	2985	101	281	343
Mikkelinseutu	738	1132	1733	177	755	1318	Simo-Kuivaniemi	355	817	1143	118	486	502
Kotka	584	833	1116	440	1079	2205	Sodankylä	1058	1964	2585	23	164	148
Vassa	430	348	460	271	1182	2049	Posio	1127	2425	3641	33	217	207
Rovaniemi	688	967	1208	190	921	1631	Ähtärinjärvi	743	897	1096	122	755	1284
Lappeenranta	738	1034	1268	253	868	1319	Kristiinanseutu	413	733	1171	208	808	1410
Hämeenlinnanseutu	674	865	1184	367	1179	1859	Tuusula	1043	1560	2385	263	779	568
Kajaani	836	1273	1883	116	843	1260	Inari-Utsjoki	1227	1835	2319	28	163	124
Kokkolanseutu	504	771	1313	272	1028	1693	Nurmijärvi	1076	1663	2166	291	790	743
Vantaa	855	895	1787	646	1465	1852	Lamminseutu	932	1454	2022	109	405	387
Kemi	782	1208	1637	266	612	1000	Kangasniemi	974	1896	2399	47	344	277
$\overline{X}$	621	814	1241	367	1133	1831	$\overline{X}$	860	1529	2142	111	449	542
SD	141	294	400	195	359	513	SD	246	452	619	90	258	448

TABLE 2. Per Cent Office-based Physician Visits by Type of Therapeutic Services Rendered for Specified Race/Ethnic Group: United States, 1979

Therapeutic Services Rendered	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent†	100.0	100.0	100.0	100.0
None	19.8	20.2	15.5	18.0
drugs (prescription and non-prescription)	51.3	50.1	62.5	54.5
Injection	9.6	9.4	11.3	13.1
Immunization/ desensitization	5.2	5.4	3.2	6.3
Diet counseling	6.0	5.8	7.7	6.9
Family planning	1.4	1.4	2.3	1.8
Medical counseling	22.2	22.3	22.1	19.1
Physiotherapy	3.1	3.0	3.7	2.9
Office surgery	7.4	7.6	4.8	7.6
Psychotherapy/therapeutic listening	4.4	4.6	3.1	2.5
Other therapy	3.5	3.6	2.1	2.8

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the United States).

† Per cent do not add to total because more than one service may be rendered per physician visit.

problem visits was 1.2 for Asian/Pacific Americans and 1.6 for white Americans, a difference which is not statistically significant.† Likewise, in terms of the kinds of diagnostic and therapeutic services rendered by the sampled physicians, so significant differences across specified race/ethnic groups were observed (Tables 2 and 3). Thus, Table 2 shows that the proportion of Asians and Pacific Islanders who received immunizations or desensitizations (6 per cent) was not significantly different from that observed for white Americans (5.4 per cent). Despite the stereotyped belief that Asians have a strong sense of modesty, an equal proportion of Asian and Pacific Islander women (5 per cent) received Pap tests as did white and black Americans (Table 3). Insofar as blood pressure check-ups are concerned, significantly more black Americans (44.5 per cent) had actually received this diagnostic service than did Asian/Pacific Islanders (33 per cent) or white Americans (35 per cent).

There are, however, significant differences between the specified racial/ethnic groups insofar as the age distributions of the patients are concerned (Table 4). For Asian/Pacific Americans, as much as 30.4 per cent (roughly 1.7 million divided by

---

† The significance of the difference between two percentages was tested using the t-test with a critical value of 1.96 (0.05 level of significance) according to the following formula:

$$t = \frac{p_x - p_w}{\sqrt{s^2 p_x + s^2 p_w}}$$

Since NAMCS is based on a complex sampling design, standard errors cannot be developed according to formulas based on the assumption of random sampling. The half-sample replication procedure was used to calculate standard errors of estimates in the survey.[5,6] Approximate standard errors of percentages based on 1979 aggregate estimates may be calculated by the following formula:

$$SE_p = \sqrt{\frac{30.96199 \times p \times (1 - p)}{X}}$$

TABLE 3. Per Cent Office-based Physician Visits by Type of Diagnostic Services Rendered for Specified Race/Ethnic Group: United States, 1979

Diagnostic Services Rendered	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent	100.0	100.0	100.0	100.0
None	10.2	10.6	6.0	8.2
Limited exam/history	63.0	63.0	63.1	64.8
General exam/history	16.8	16.6	18.5	20.3
Pap test	4.9	5.0	4.7	5.0
Clinical lab test	23.2	23.0	25.3	21.8
X-ray	8.2	8.2	8.6	5.2
Blood pressure check	36.0	35.3	44.5	33.2
EKG	2.7	2.7	3.1	0.5
Vision test	6.0	6.2	4.5	6.6
Endoscopy	1.3	1.4	0.8	1.8
Mental status exam	1.5	1.5	1.3	0.4
Other diagnostic service	3.5	3.6	2.5	3.9

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the coterminous United States).

† Per cent do not add up to total because more than one service may be rendered per physician visit.

5,559,524) of estimated visits to physician's office were made by children under 15 years of age, which is significantly different from that observed for white or black Americans (18.2 per cent and 17.1 per cent, respectively). Using the population estimates obtained from the National Health Interview Survey by N.C.H.S. in 1979, we have calculated the visit rate for the different racial/ethnic groups. The findings are shown also in Table 4. First, compared with white or black Americans, Asian/Pacific Americans have the lowest visit rate to office-based physicians. Second, while

TABLE 4. Estimates and Rates of Visits to Office-based Physicians by Age for Specified Race/Ethnic Group: United States, 1979

	All Races		White		Black		Asian/Pacific	
Age Group	Visit Estimates	Visit Rate*	Visit Estimates	Visit Rate*	Visit Estimates	Visit Rate*	Visit Estimates	Visit Rate*
All ages	556,313,431	2.58	502,926,839	2.70	46,789,200	1.84	5,559,524	1.55
Under 15 years	101,352,298	2.02	91,267,680	2.21	8,015,492	1.03	1,690,800	1.98
15–24 years	82,289,782	2.05	74,295,285	2.18	7,209,758	1.38	675,243	0.99
25–44 years	151,713,912	2.59	133,337,485	2.61	16,485,828	2.58	1,590,530	1.49
45–64 years	128,594,299	2.96	117,859,164	3.04	9,550,199	2.35	1,061,339	1.51
65 years and over	92,363,140	3.96	86,167,225	4.08	5,527,923	2.71	541,612	1.94

* Based on estimates of the civilian, noninstitutionalized population of the United States excluding Hawaii and Alaska, obtained from the Health Interview Survey, 1979. In the HIS Program, race for black and white Americans was based on interviewer's report. Race of Asian/Pacific Americans was based on respondent's self-report.

the visit rate for white and black Americans shows an almost linear increase with age, that for Asian/Pacific Americans displays a more U-shaped relationship with age. Third, within-group comparison indicates that among Asian/Pacific Americans, in contrast to the majority of Americans, the youngest age group have the highest visit rate.

In terms of medical specialty, about one-third of the visits for both white and Asian/Pacific Americans were made to the offices of General and Family Practitioners. However, significantly more visits by Asians and Pacific Islanders (19 per cent) compared with white majority (10 per cent) or black minority (10.6 per cent), were made to the offices of pediatricians. These data are, of course, consistent with the earlier finding that a sizable proportion of the physician visits by Asian/Pacific Islanders were made by patients under 15 years of age. Dermatology is another specialty where a statistically significant difference exists between Asians and white Americans. In addition, a significantly smaller percentage of the visits by Asians and Pacific Islanders vis-a-vis white majority were made to the office of a general or specialty surgeon or a psychiatrist (Table 5). The cultural underpinnings of these findings will be discussed in the following section of this paper.

Table 6 presents data on physician visits by major ICD-9 Diagnostic Codes[7] for specified race/ethnic groups. Only 1.9 per cent of the visits made by Asian/Pacific Americans compared with 4.5 per cent for the white majority were for mental disorders. The difference is statistically significant. In contrast, significantly more visits made by Asian/Pacific Islanders (10.2 per cent) than by white Americans (5.3 per cent) were for problems related to diseases of the skin and subcutaneous tissue. Examination of the more detailed ICD-9 codes reveals a clue to the underlying reasons for this peculiar "skin problem" among Asian/Pacific Islanders. Contact dermatitis and other eczemas, as well as diseases of sebaceous glands are two diagnoses in which the per cent visits are proportionally higher among Asian/Pacific Americans, compared with white and

TABLE 5. Per Cent Office Visits by Physician Specialty for Specified Race/Ethnic Group: United States, 1979

Medical Specialty	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent	100.0	100.0	100.0	100.0
General and family practice	34.2	33.4	42.3	32.0
Internal medicine	12.0	12.3	9.3	9.3
Pediatrics	10.5	10.3	10.6	19.3
Obstetrics and gynecology	9.1	9.0	10.2	13.5
Dermatology	3.2	3.3	1.2	7.1
Other medical specialities	3.8	4.0	2.7	1.9
General surgery	6.1	6.1	6.5	3.8
Other surgical specialty	16.0	16.3	13.4	11.7
Psychiatry	3.1	3.2	1.7	0.6
Other	2.1	2.1	2.5	0.8

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the coterminous United States).

black Americans. Since these diagnoses include diaper rash and acne, their relative preponderance is most likely an artifact of the overrepresentation of children under 15 years old among Asians and Pacific Islanders.

Table 6 also shows that 23 per cent of the Asian or Pacific American visits compared with only 16 per cent for the white majority were made for the category "factors influencing health status and contact with health services" (which used to be called "special conditions and examinations without sickness"). This diagnostic group includes general medical examinations, routine normal pregnancy (prenatal) examinations, and supervision of healthy children (well-baby examinations), which together constitute the largest proportion of visits for specific diagnoses regardless of race. Of further interest is the significantly smaller proportion of injury and poisoning (which was called "accidents, poisoning, and violence" in the ICD-8) observed for Asian and Pacific Islanders (5.3 per cent) compared with white (9.3 per cent) or black (9.6 per cent) Americans. The implications of these findings are explored below.

Cultural Diversity Reflected in NAMCS Data

In many ways, the results of our analyses are both heartening and disheartening. On one hand, the visit rate for Asian/Pacific Americans obtained from the NAMCS data is rather low in every age group except for those under 15 years, compared with that found for white and black Americans. A perusal of the National Health Interview Survey data at NCHS lends an interesting interpretation to the NAMCS finding. From 1976 through 1978, visits to office-based physicians on the average accounted for 66.5 per cent of the total annual encounters by Asian/Pacific Islanders with sources of medical care, compared with 69.3 per cent for white Americans (NCHS, unpublished HIS data).‡ In contrast, visits to emergency rooms and/or outpatient clinics represented as much as 18.8 per cent of the total visits by Asian/Pacific Americans, compared with 11.7 per cent for white Americans. In terms of the rate of visits to emergency rooms and/or outpatient clinics per 1,000 population, Asian/Pacific Americans had a higher rate—798.1§ (annual average from 1976–78 pooled data)—than white Americans in 1978 (566.0§) but a lower rate than black Americans (1,188.8§ for 1978). These NHIS findings suggest that the NAMCS data give only a partial picture of the use of health services by Asian/Pacific Americans. It is most likely that language and cultural barriers have prevented many immigrant Asians and Pacific Islanders from having a regular "family" doctor, as had been reported earlier in a few localized studies.[8,9] Lack of health insurance, ignorance about the service delivery system in the United States, and a sizable number of unattached adults and elderly in the Asian populations are three other possible factors in the heavy reliance of immigrants on emergency room and/or outpatient clinics for medical care.

On the other hand, at least one reason for optimism exists. Taken at face value, the NAMCS data fail to show gross disparities—across specified race/ethnic groups—in the kinds of diagnostic and therapeutic services rendered by the sampled physicians to their patients. Instead, some interesting cultural diversity in medical care is reflected in the NAMCS data. For instance, within the Asian/Pacific Islander population, children under 15 years of age have the highest rate of visit to

---

‡ Prior to 1977, racial breakdowns in the HIS tabulations did not routinely allow for the identification of black Americans separately.

§ The visit rate is age-adjusted by the direct method to the 1970 civilian noninstitutionalized population, using four age intervals: Under 17 years, 17–44 years, 45–64 years, and 65 and over.

TABLE 6. Per Cent Office-based Physician Visits for Specified Race by Principal Diagnoses and ICDA Code: United States, 1979

Principal Diagnosis and ICD-9-CM Codes*		Total* U.S.	White	Black	Asian/ Pacific Americans
All classes					
estimate of visits		556,313,431	502,926,839	46,789,200	5,559,524
per cent		100.0	100.0	100.0	100.0
Infectious and Parasitic Diseases	(001–139)	3.5	3.5	3.4	4.1
Neoplasms	(140–239)	2.6	2.7	1.6	1.4
Endocrine, nutritional, and metabolic diseases, and immunity disorders	(240–279)	4.1	4.0	5.1	3.2
Mental disorders	(290–319)	4.4	4.5	4.1	1.9
Diseases of the nervous system and sense organs	(320–389)	9.1	9.4	6.3	9.1
Diseases of the circulatory system	(390–459)	8.9	8.8	10.6	7.3
Diseases of the respiratory system	(460–519)	13.2	13.1	13.6	15.1
Diseases of the digestive system	(520–579)	4.4	4.5	3.9	6.6
Diseases of the genitourinary system	(580–629)	6.6	6.3	9.9	5.0
Diseases of skin and subcutaneous tissue	(680–709)	5.2	5.3	3.8	10.2
Diseases of the musculoskeletal system and connective tissue	(710–739)	6.7	6.5	8.3	4.0
Symptoms, signs, and ill-defined conditions	(780–799)	3.1	3.1	2.9	1.8
Injury and poisoning	(800–999)	9.3	9.3	9.6	5.3
Factors influencing health status and contact with health services	(V01–V82)	15.8	15.9	14.2	23.3
Other diagnoses		1.5	1.5	1.6	0.6
None or unknown		1.6	1.6	1.1	1.1

* Based on the Ninth Revision, International Classification of Diseases, Clinical Modification, September 1980.

† Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the coterminous United States).

‡ Includes diseases of the blood and blood-forming organs, complications of pregnancy, childbirth and the puerperium, congenital anomalies, certain causes of perinatal morbidity and mortality, blank diagnosis; noncodable diagnosis; and illegible diagnosis.

office-based physicians, compared with the other age groups. We suspect that this may be due to the presence of a sizable proportion of persons in the under 15 age group because Asian/Pacific Americans are, by and large, still very much an immigrant population. This is borne out by an examination of the U.S. Census data on the proportion of foreign-born persons for different race/ethnic groups. For instance, while only five per cent of the white population in 1970 was born outside of the

United States, the lowest percentage of foreign-born persons within the five *major* Asian American subgroups[‖]—21 per cent—is found among Japanese Americans.[10] An inspection of thȩ annual reports prepared by the Immigration and Naturalization Service[11] reveals that the influx from the Philippines, Korea, China (including Hong Kong and Taiwan), and IndoChina (Vietnam, Laos, Cambodia)—in contrast to Japan—has increased dramatically as a result of the relaxation of U.S. immigration restrictions in 1965. Hence, the foreign-born segment of the Chinese population (47 per cent as of 1970 Census) and that of the Filipino population (53 per cent as of 1970) is likely to increase dramatically when the 1980 Census tabulations are released within the next two years. Although corresponding figures for Koreans and IndoChinese are not available from the 1970 Census, the immigration records provide evidence that the majority of these two Asian subgroups will be foreign-born by virtue of their sudden surge in the 1970s. Some researchers have estimated that it took the IndoChinese less than five years to reach the population size in the U.S. comparable to that achieved by the Chinese in over 100 years of piecemeal immigration—a conjecture which remains to be tested against the 1980 Census results. Thus, we can expect that the majority of the Asian Americans will continue to retain many of their traditional values and health practices from the old countries. Accordingly, health is viewed as a state of equilibrium ("homeostasis") between man, society, and the cosmic forces of the universe. Maintaining a balance between the "hot" and "cold" elements of the body is the cornerstone of good health. Disease is a disturbance of that relationship. Blood is a source of human vitality and difficult to replenish. Consequently, surgery is avoided to the extent possible since it increases the risk of losing one's blood. These concepts, of course, are so deeply rooted in the traditional cultures of developing societies that they crisscross ethnic boundries. For various Asian ethnic groups, however, some of the similarities in herbal prescriptions (e.g., the use of Ginseng for restoring vitality or stress endurance) and behavioral proscriptions (e.g., never sleeping with wet hair; "doing the month" after delivering a baby) are rather striking.[12–16] Taking into consideration the major composition of the Asian/Pacific American population, i.e., Chinese, Japanese, Korean, IndoChinese, and Filipino, it is apparent that historically at least four of these groups have been heavily influenced by China—in language, religion, philosophy, social structure, and medicine. Such pervasive influence was made possible by the diffusion and use of Chinese characters as the primary vehicle of written communication and the medium of instruction among the literati in ancient Japan (5th to 19th century), Korea (108 B.C.–1886), and Vietnam (111 B.C.–16th century). Although the Filipinos had their unique alphabet system, the fact that Chinese presence in the Philippines antedates the Spanish Conquest (began in the 16th century) by at least 500 years, also suggests the possibility of Chinese influence in Filipino society—though not to the same extent as that found in other countries neighboring China. Therefore, despite the apparent cultural diversity among Asians, unifying threads of social values and health beliefs are interwoven into the indigenous cultures. The persistence of these health-related beliefs and practices in Asian American subcultures is evident from a few studies in the United States based on limited samples and sometimes impressionistic observations.[17–19] Obviously, precise empirical support for the prevalence and sociodemographic variations of these ethnic practices are lacking.

[‖] Chinese, Japanese, Filipino, Koreans, and Vietnamese.

In the NAMCS data, demographic factors probably played an important role in the high physician-visit rate of Asian/Pacific Americans for "special conditions and examinations *without* sickness," while cultural factors may be salient in their unusually low proportion of injury and poisoning (or what was called "accidents, poisoning, and violence" in the ICD-8). The former is essentially *preventive* health care, while the latter, though usually unanticipated, is *preventable* in most cases. Thus, the juxtaposition of these two findings would tend to suggest that although proportionally fewer Asian/Pacific Americans rely upon office-based physicians for their regular source of medical care, many of those who do visit physicians are apparently prevention-oriented. It is further possible that this consciousness for the importance of prevention may extend beyond the supervision of infants.

As early as 1966, Kurokawa[20,21] noted that "Oriental" children in California had had far fewer accidents than Caucasian children. She hypothesized that this may be due to the peculiar family structure found among Asians (primarily Japanese and Chinese) where adults other than the child's parents are not only present, but also participate in the supervision and discipline of the young. In contrast, Caucasian families are mostly nuclear in structure, and parental rights are seldom fully shared with other adult members, even when they are present. To further prove her point, Kurokawa interviewed the mothers of 151 Oriental children who were enrolled in the Kaiser prepaid medical plan in California. She found that acculturated (i.e., "Americanized") Oriental children had more accidents than the unacculturated, because the former were more likely to be independent and venturesome which, in turn, exposed them to potential hazards, whereas the latter tended to be "over-supervised" by American standard. Findings from other independent studies indicate that traditional Asian cultures indeed reinforce dependency behavior in the child, which restricts his ability to explore the unfamiliar all by himself.[22–24] American child socialization, on the other hand, places great emphasis on independence training.[25,26] Hence, it is possible that, relatively speaking, American child socialization exposes the young to greater risks of injury due to accidents, poisoning, and violence. Nevertheless, the small sample size in Kurokawa's study makes broad generalizations from her findings difficult. Regardless of this limitation, her work does provide some food for thought and demonstrate the dire need for more systematic studies on the family structure and health practices of Asian/Pacific Americans.

Another interesting cultural pattern manifested in the National Ambulatory Medical Care Survey data is the significantly small percentage of visits made by Asian/Pacific Americans to the office of a psychiatrist. Moreover, regardless of the medical specialty of the physician they visited, only a very small percentage of Asian/Pacific American patients received the principal diagnosis of "mental disorders." Thus, the NAMCS data lend confidence to previous observations made by practicing psychiatrists and clinical psychologists. To illustrate, in a study of 17 community mental health services in the greater Seattle area, Sue[27] found that over a period of three years, only 3.1 per cent of the Asian American clients saw a psychiatrist at intake, and not one Asian saw a psychiatrist during therapy. A number of researchers believe that cultural factors play an important role in the asian reluctance to consult with psychiatrists, even when they admit to having psychological problems. For the most part, Asians and Pacific Islanders appear to have a tendency to somaticize their psychiatric symptoms.[28] Hence, the consensus among practitioners is that members of this minority group generally perceive "talk" therapy to be ineffective.[29,30] Instead, self-medication, such as herbal tonics, are the expected treatment even for mental problems, thereby causing an extensive delay in help-seeking until the disorder has clearly become unbearable or unmanageable. Lin and Lin[31] reported from their experience that the delay among Chinese patients in seeking psychiatric help can take as long as 25 years. Intrafamilial resources are first utilized before outside help is sought and, even then, respected intermediaries who have the family's trust are consulted before assistance is sought from a psychiatric professional who is a total stranger to the family. These findings, consistent as they are, are not based on large-scale epidemiologic studies. Hence, until prevalence and incidence data on both "treated" and "untreated" mental disorders among Asian/Pacific Americans are available, the extremely low percentage of visits to psychiatrists by this growing minority must be interpreted with caution.

Conclusion

Taking into consideration the limitations of the data collected from the National Ambulatory Medical Care Survey conducted by the National Center for Health Statistics, we have presented the major findings on physician visits by Asian/Pacific Americans. On one hand, no significant difference was observed between Asian/Pacific Islanders and the white majority with regard to the patient's sex, return-visit ratio, or the kinds of diagnostic and therapeutic services rendered by the sampled physicians. On the other hand, some interesting cultural differences are discernible from the patterns of visits to office-based physicians reported for Asian/Pacific Islanders. Children, followed by the elderly, appear to be the major utilizers of ambulatory medical care among Asian/Pacific Americans. However, in almost every age group, the visit rate to office-based physicians is lowest for Asian/Pacific Americans, compared with either black or white Americans. When compared with white Americans, a significantly small percentage of Asians or Pacific Islanders had visited a physician for injury and poisoning; and a substantial proportion of their visits were apparently made for preventive care—general medical examinations, routine normal pregnancy examinations, and supervision of healthy children. Significantly few visits were made by Asian and Pacific Americans to the office of a surgeon or a psychiatrist. A cultural resistance to the utilization of these two types of medical specialty is suggested.

Acknowledgments

The authors thank their colleagues Edward Bacon (Director of the Division of Health Care Statistics, N.C.H.S.), Joel Kleinman (Director of the Division of Analysis, N.C.H.S.), William T. Liu (Director of the Pacific/Asian American Mental Health Research Center and Professor of Sociology, University of Illinois at Chicago Circle), and two anonymous reviewers for *Medical Care*, for their constructive criticisms of an earlier draft of this paper. Another colleague, Thomas Drury (Acting Chief, Health Status Measurement Branch), was most helpful in sharing some of the unpublished HIS data referred to in this paper.

References

1. Bureau of the Census. 1980 Census Population Totals for Racial and Spanish Origin Groups in U.S. Announced by Census Bureau. Washington, D.C.: United States Department of Commerce, Public Information Office, Feb. 23, 1981.

2. National Center for Health Statistics. 1979 summary ambulatory medical care survey. Advancedata 1981;66:1.

3. Yu E. A sociodemographic profile of Asian Americans, 1820–1976. Paper presented at the Ethnic Resources Institute of Rosary College, River Forest, Illinois, October 19–20, 1979.

4. Yu E, Drury T, Liu WT. Estimating the health status of Asian/Pacific Americans: conceptual and methodological problems. Papers prepared for the Proceedings of the Annual Meeting of the American Statistical Association, Detroit, Michigan, August 1981.

5. McCarthy PJ. Replication: an approach to the

analysis of data from complex surveys. Vital and Health Statistics. Series 2, No. 14. DHEW Pub. No. (HSM) 73-1269. National Center for Health Statistics, Health Services and Mental Health Administration. Washington: U.S. Government Printing Office, April 1966.

6. McCarthy PJ. Pseudoreplication: further evaluation and application of the balanced half-sample technique. Vital and Health Statistics, Series 2, No. 31. DHEW No. (HSM) 73-1270. National Center for Health Statistics, Health Services and Mental Health Administration. Washington: U.S. Government Printing Office, January 1969.

7. World Health Organization. International Classification of Diseases. Manual of the international statistical classification of diseases, injuries, and causes of death, vol I. Geneva: W.H.O., 1980.

8. Li F, Schlief NY, Chang CJ, Gaw AC. Health care fo the Chinese community in Boston. Am J Public Health 1972; 62:636.

9. Chinatown, U.S.A., 1970. California Health 1970;2:1.

10. Sung BL. Chinese american manpower and employment. Washington, D.C.: U.S. Department of Labor Manpower Administration, 1975.

11. Immigration and Naturalization Service. Immigration and Naturalization Reports, 1943–1979. Washington, D.C.: U.S. Government Printing Office.

12. McKenzie JL, Chrisman NJ. Healing herbs, gods, and magic: folk health beliefs among Filipino-Americans. Nurs Outlook 1977;25:326.

13. Gould-Martin K. Hot cold clean poison and dirt: Chinese folk medical categories. Soc Sci Med 1978;12:39.

14. Popov IM, Goldwag WJ. A review of the properties and clinical effects of ginseng. Am J Chinese Med 1973;1:263.

15. Sich D. Traditional concepts and customs on pregnancy, birth and post partum period in rural Korea. Soc Sci Med 1981;15B:65.

16. Pillsbury, BLK. "Doing the Month": confinement and convalescence of Chinese woman after childbirth. Soc Sci Med 1978;12:11.

17. Ling S, King J, Leung V. Diet, growth, and cultural food habits in Chinese-American infants. Am J Chinese Med 1975;3:125.

18. Campbell T, Chang B. Health care of the Chinese in America. Nurs Outlook 1973;21:245.

19. Chan CW, Chang JK. The role of Chinese medicine in New York City's Chinatown. Am J Chinese Med 1976;4:129.

20. Kurokawa M. Family solidarity, social change, and childhood accidents. J Marriage Fam 1966;28:498.

21. Kurokawa M. Acculturation and childhood accidents among Chinese and Japanese Americans. Genet Psychol Monogr 1969;79:89.

22. Yu E, Liu WT. Fertility and Kinship in the Philippines. Notre Dame: University of Notre Dame Press, 1980.

23. Young NF. Independence training from a cross-cultural perspective. Am Anthropologist 1972;74(3):629.

24. Vogel E, Vogel S. Family security, personal immaturity, and emotional health in a Japanese sample. J Marriage Fam 1961;23:161.

25. Winterbottom M. The Relation of Childhood Training in Independence to Achievement Motivation. M. A. thesis, University of Michigan, 1953.

26. Kriger SF, Kroes WH. Child rearing attitudes of Chinese, Jewish, and Protestant mothers. J Soc Psychol 1972;86:205.

27. Sue S. Community mental health services to minority groups. Some optimism, some pessimism. Am Psychologist 1977;32:616.

28. Tseng WS, McDermott JF Jr, Maretzki TW, eds. People and Cultures in Hawaii: An Introduction for Mental Health Workers. Honolulu: Department of Psychiatry, University of Hawaii School of Medicine, 1974.

29. Chien C-P, Yamamoto J. Effective psychotherapy for the Asian American and Pacific Islander patients. In: Acosta F, et al., eds. Psychotherapy for Low Income and Minority Patients. Prepublication copy, 1981.

30. Tung TM. The IndoChineses Refugees as Patients. Refugee Resettlement Exchange Project. Washington, D.C.: American Public Welfare Association, 1980.

31. Lin T-Y, Lin M-C. Service delivery issues in Asian–North American communities. Am J Psychiatr 1978;135:454.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Drug Product Selection in Illinois: The First Year's Experience

JESSE E. STEWART, PH.D., PAUL G. GRUSSING, PH.D., AND ANAL A. PUROHIT, PH.D.

This paper reports an evaluation of Illinois' drug product selection (DPS) law. The results show that while consumers saved an average of $1.47 whenever DPS occurred in the study period, the prescription format mandated in the Illinois law appears to be a hindrance to unambiguous communication from prescribers. In addition, because of the positive formulary approach adopted by Illinois, the potential for DPS behavior was limited to only 20.8 per cent of the prescriptions audited in the study and ultimately, DPS occurred in 12.6 per cent of this sample. It also is reported that DPS was more likely to occur in urban community pharmacies and that consumers saved more money when DPS occurred in this environment, but prices for prescriptions dispensed as written were also higher in this environment. Finally, it was found that prescription prices were affected by the type of pharmacy patronized, but were not affected by the urban or rural location of pharmacy.

ALTHOUGH DRUGS PRESENTLY account for less than 10 per cent of the total cost of medical care in the United States, they have been identified as representing an area of potentially substantial savings to consumers, for a variety of reasons. Proponents of prescribing and dispensing of drugs by generic (rather than brand) name, such as the Federal Trade Commission and the Department of Health and Human Services through the Food and Drug Administration, argue that a considerable portion of consumer expenditures on prescription drugs could be saved if the practice of selecting less expensive generic drug products was encouraged.[1,2]

For the 20 years before 1970, virtually all states had laws which specifically forbade the pharmacist to dispense a drug manufactured or distributed by any company other than the brand-name drug called for on the prescription. Only if the prescription was written generically could the pharmacist dispense any manufacturer's equivalent product in the pharmacy. In the early 1970s, the states began repealing these so-called antisubstitution laws, however, replacing them with a variety of laws aimed at saving consumers' money.

These drug product selection (DPS) laws allow the pharmacist, usually in specified instances, to dispense a less expensive generically equivalent drug product in lieu of the branded product originally prescribed. A variety of mandatory

---

From the Department of Pharmacy Practice, University of Illinois, College of Pharmacy, Chicago, Illinois.

Supported in part by the Illinois Department of Public Health, Springfield, Illinois. Data coding and computerization was performed by the Survey Research Laboratory, University of Illinois, Chicago Circle Campus. The paper represents part of a larger research effort which is being reported in several independent communications. Special thanks are expressed to Mr. Ronald Gottrich who as a Graduate Research Associate was invaluable in the successful completion of this project.

Address for reprints: Jesse E. Stewart, Ph.D., University of Illinois, College of Pharmacy, 833 South Wood Street, Room 142, Chicago, IL 60612.

0025-7079/82/0800/0831/$01.40 © J. B. Lippincott Co.

and permissive clauses in the various laws have produced an environment in which the laws can basically be classified in three categories. The first and largest category includes those laws which do not require a formulary,* but simply allow the pharmacist to dispense a less expensive, generically equivalent product, unless prohibited by the prescriber or the patient. The second category, representing almost as many states, includes those laws which mandate the development of a positive formulary. Thus, when pharmacists select drug products, they can only do so within the limitations of a formulary, which generally lists those multiple-source drugs judged to be equivalent (either therapeutically or chemically, depending upon state law) by a committee of experts. The last category includes those states enacting laws that require the development of a negative formulary. In these states, the pharmacist may use drug product selection for any drug unless it is listed in this formulary or prohibited by the prescriber. The DPS laws in 28 states now require the development of a positive or negative formulary. A variety of additional clauses, such as prescription signature format, prevention of DPS by the prescriber, cost savings pass-on requirements and patient consent requirements add to the variability of laws from state to state and have been accurately summarized elsewhere.[3]

Drug product selection laws have now been enacted in almost every state, but evidence of how effective the various laws are is still limited. Only a few studies have evaluated the effectiveness of the DPS laws through the use of actual prescription data. While some[4] might argue that shopping surveys represent the most appropriate methodologic approach, the expense involved in using this approach generally has forced researchers to review only a small number of drugs in a limited number of pharmacies. Thus such results are limited in their applicability to other geographic areas or even to other drugs or drug classes.

The first major evaluation of DPS legislation was conducted by Goldberg and colleagues in Michigan, using Wisconsin as a control state.[5–7] The Michigan law, which is permissive by not mandating DPS, does not use a formulary and allows prescribers to prohibit DPS by stating so on the prescription. By examining prescription data for the year prior to enactment of Michigan's law and for the subsequent 2 years, they found that significant savings occurred when DPS took place ($1.23 per prescription in year 2) as well as when prescriptions were written generically. The savings were less dramatic, however, when the prescription initially was written generically. They also reported that the rate of DPS did not increase substantially from the first year of the law to the second, and was found to be only 1.50 per cent of all prescriptions for multiple source products.

A second major study recently was reported by Vuturo and colleagues following their evaluation of Florida's DPS law.[8] A unique feature of the Florida law requires the state to develop a negative formulary and each pharmacy to develop and display its own positive formulary for use when selecting drug products. The researchers reported that an average savings of $1.92 was achieved each time DPS occurred in their study sample, and despite the formulary limitations of the Florida law, the DPS rate of 6.2 per cent for multiple-source prescriptions compared favorably with the 1.5 per cent reported in Michigan and the 2.7 per cent reported in Wisconsin.

The present study represents the third major study examining the effectiveness of a state's DPS law that used actual prescription data collected on a statewide basis and was not restricted to a limited number of drugs for comparison. On July 1, 1978, the Illinois Drug Product Selection law became effective and incorporated a number of features found in the laws of other states. The law amended the Illinois Food, Drug and Cosmetic Act to allow DPS if the product is listed in a positive formulary to be developed, maintained and issued by the Illinois Department of Public Health. The remaining features of Illinois' law were enacted as amendments to the Illinois Pharmacy Practice Act and are highlighted below:

---

* A list of drug entities showing pharmaceutically or therapeutically equivalent products within each entity. A formulary includes both branded products and those marketed only under a generic name.

1. The prescription blank must contain two signature lines, one indicating "may substitute," the other indicating "may not substitute," and two accompanying check boxes.
2. DPS may occur when the "may substitute" option is properly signed and checked or when the prescriber fails to properly check and sign beside one of the alternatives.
3. When presented with a prescription where drug product selection can be used, the pharmacist may inform the patient of the availability of a less expensive product and the price of the alternate product.
4. The person for whom the prescription is written, a spouse, or a guardian can provide oral permission to the pharmacist allowing DPS.
5. The alternate product must be less expensive than the product that was originally called for in the prescription.
6. When DPS occurs, the pharmacist must record the name and address or other identification of the manufacturer of the product ultimately dispensed.
7. The act of DPS by a pharmacist will not constitute evidence of negligence if the alternate drug is included in the Illinois Drug Formulary.

These seven areas represent the legislative intent as well as the mechanics of the Illinois law. In addition, a unique feature of the law states that "the Department of Public Health shall have the duty of monitoring the cost savings effected by substitution . . . and shall annually report to the legislature as to the cost savings being achieved." A study was, thus, designed to answer this legislative mandate as well as to examine several other important features of the law in order to provide evidence of the law's effectiveness. That portion of the study described here had three primary objectives:

1. Measure the rate of DPS in Illinois according to pharmacy type, payment and geographical variables.
2. Identify certain factors which preclude or promote DPS behavior by physicians and pharmacists.
3. Determine cost savings to consumers as a result of the DPS law.

Methodology

To answer the evaluation questions posited in this research, the prescription was identified as the basic element or primary unit of investigation. The study population was thus defined as all new prescriptions dispensed between May 1, 1978, and June 30, 1979, in Illinois pharmacies that were subject to the Drug Product Selection law.†

A two-stage sampling design was utilized. The first stage sampling involved selection of pharmacies and identified geographic location and type of pharmacy as important selection variables. All licensed pharmacies in the state were divided into urban and rural categories as defined by Standard Metropolitan Statistical Areas. Next, all pharmacies were classified as either community, chain (four or more locations), or hospitals providing outpatient pharmacy services using acceptable definitions. Clinic pharmacies were combined with community pharmacies and a total of 227 other pharmacies, such as nursing home pharmacies and nuclear pharmacies, not affected by the law, were excluded from the study.

---

† The 14-month study period represented the first 12 months after the law had been enacted (July 1, 1978–1979) plus the two-month period prior to enactment for data baseline purposes. All data reported are based on the 12-month study period only.

Initially, 60 pharmacies were selected systematically with a random start from the list of licensed pharmacies generated previously. The sample size for each of the six strata (community, chain and hospital pharmacies versus urban or rural location) was selected proportional to the number of pharmacies in each stratum. The rural chain cell and hospital outpatient cell were oversampled choosing six pharmacies in each cell since initial selection of one pharmacy in each cell was deemed to be too small for comparison purposes.‡ In addition, because of the disproportionately small number of hospitals providing outpatient pharmacy services, the urban and rural hospital cells were collapsed. Thus, data were ultimately collected from 70 systematically selected pharmacies representing the three types of pharmacies in both urban and rural environments, with the hospital cell representing both environments.

The second stage of the sample selection involved identification of new prescriptions dispensed during the 14-month study period. Due to the expected low rate of DPS, a large number of prescriptions needed to be examined in order to answer the evaluation questions. However, because most prescriptions were expected to be dispensed as written (DAW), it was not necessary to generate a proportionately large sample of prescriptions to determine the prescribing and dispensing characteristics of physicians and pharmacists. To accommodate these extremes and to facilitate comparison with previous studies, the methodology of auditing and reviewing prescriptions used in previous studies was adopted.[7]

On the average, 200 clusters of 10 prescriptions were surveyed from the 70 pharmacies. The first prescription of the cluster of 10 was an "audited" prescription and the remaining nine prescriptions of the cluster were considered "reviewed" prescriptions. The purpose of the audited prescription was to collect detailed data about the prescription. The purpose of the reviewed prescription was to enlarge the sample of drug-product-selected prescriptions without necessarily enlarging the sample of DAW prescriptions, since the DPS rate, the unit of study, was expected to be low. For the reviewed prescriptions, such detailed data were collected only if drug product selection was used for the prescription. Using this methodology, a theoretical total of 140,000 prescriptions would be examined visually (200 × 1 × 70 audited prescriptions + 200 × 9 × 70 reviewed prescriptions).

To select prescriptions proportional to pharmacy prescription volume, each of the 70 selected pharmacies was telephoned and asked to provide their prescription volume for the 14-month study period.§ Prescription volumes then were determined separately for each of the five cells, and selection intervals for the clusters of 10 prescriptions were determined for each cell independently. The sample of prescription clusters were then selected proportional to the prescription volume in each pharmacy. Thus, larger volume pharmacies had more prescriptions audited and reviewed than did smaller pharmacies (i.e., more clusters selected).

Data were collected by the investigators as well as by graduate students in pharmacy administration. A one-day training session was completed by all data collectors which included actual use of the data collection form. On the average, the error rate of data collectors was less than three per cent, and subsequent to actual data collection, each individual was retrained in those areas where errors had occurred. Thus, the error rate in the field was expected to be even lower.

---

‡ Appropriate weighting procedures were subsequently employed so that the statistical analysis of data would reflect the number of pharmacies accurately representing each cell.

§ Each of the reported volumes were later verified by data collectors in each pharmacy and no volume or sampling adjustments were subsequently required.

Results and Discussion

The primary intent of DPS legislation is to decrease the amount of money that consumers must spend on prescription drugs. The following discussion will examine: 1) the potential market for DPS in Illinois, 2) the rate of DPS in Illinois, and 3) the savings occurring when the pharmacist practiced DPS.

Potential for Drug Product Selection

Drug Product Selection can only occur in Illinois if the prescription is written using a product's brand name, if that drug entity[‖] is listed as interchangeable in the *Illinois Drug Product Selection Formulary* and if both the prescriber and patient allow the use of an alternate product. While each of these factors was incorporated into the law to protect the consumer and the professional prerogative of the prescriber, each tends to limit the potential instances of DPS.

Earlier studies in Michigan[6] and Florida[8] noted that just over 50 per cent of the prescriptions examined were multiple source and, thus, potentially eligible for DPS. The remainder of the prescriptions were single source products, most of which were still under patent protection. In Illinois, even when a prescription drug is multiple source, it must be written using a

---

[‖] Drug entity—a unique drug with individual characteristics and specifications. Several equivalent brands or specific drug products may occur for the same drug entity.

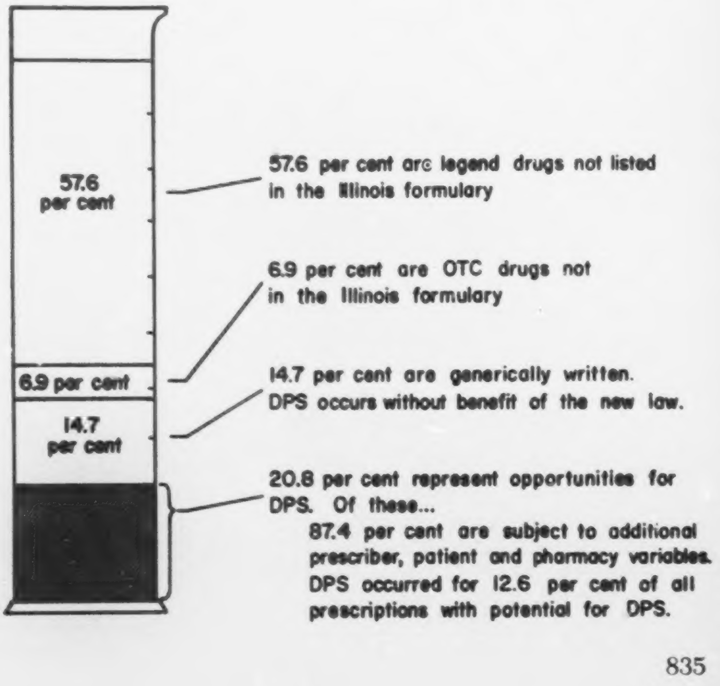


FIG. 1. Opportunities for drug product selection. Data represent audited prescriptions (N = 12,071) for the first 12 months of Illinois' DPS law.

brand name and be listed in the Illinois Formulary before DPS can be even considered. Those prescriptions written using the drug's generic name may be dispensed by the pharmacist with either a brand name product or a less expensive generically marketed product, just as occurred prior to this law.

Figure 1 graphically displays the opportunities for DPS within the confines of the Illinois law. Of the 12,071 prescriptions audited in the 12-month study period, 57.6 per cent were legend¶ drugs not listed as interchangeable in the Illinois Formulary and are referred to as nonformulary. Another 14.7 per cent (12.3 legend and 2.4 per cent over-the-counter) were written generically and were not affected by this law. Products not requiring a prescription (over-the-counter) are excluded from the Illinois law also. These represented 6.9 per cent of all audited prescriptions. Finally, it can be seen that only 20.8 per cent of all prescriptions audited in the study were written using the drug's brand name and were listed in the Illinois Formulary. This represents the potential area of opportunity for the pharmacist to practice DPS, given physician and patient consent. The rate or occurrence of DPS is most appropriately expressed based on this category alone.

Rate of DPS

An examination of those prescriptions that were legend, brand written, and listed in the formulary (20.8 per cent of all audited prescriptions) indicates that DPS occurred 12.6 per cent of the time. These results compare favorably to data reported for Florida (14.2 per cent)[8] and Vermont (14.4 per cent),[9] but are in marked contrast to the very low rate reported for Michigan (1.5 per cent).[7] A complete discussion of comparison of DPS experiences across those few states reporting such data is beyond the scope of this paper, but the reader is cautioned that differences in state laws, research methodologies and means of reporting the data prevent direct comparisons unless a thorough understanding of these differences is achieved first.

In order to examine whether certain identifiable environmental factors affect the rate of DPS in Illinois, an analysis of rates was conducted for each of the five major pharmacy types studied. Table 1 indicates that urban community pharmacies showed the highest rate of DPS (15.3 per cent) while hospital pharmacies showed the lowest rate for their outpatient dispensing activities. The high rate for urban community pharmacies is explained at least partly by higher proportional numbers of Illinois Public Aid (Medicaid) prescriptions dispensed in this environment with that program effecting de facto DPS for financial reasons. The low rate (five per cent) reported in the hospital outpatient environment reflects a high rate of generic prescribing in the environment (18.3 per cent) combined with the existence of a hospital formulary which in itself restricts prescribers to a limited number of brand name products which they learn to prescribe.

In addition to differences in DPS rates between pharmacy types, urban versus rural environments also produced notable differences. The rate of DPS for all urban pharmacies studied was 14.3 per cent while DPS was much less likely to occur in rural pharmacies (7.0 per cent rate).

While the DPS rate in Illinois is dependent upon the fact that the prescription is written using a brand name, and that the drug entity is listed in the Formulary it also is dependent upon the physician's signature and approval format. As mentioned, the Illinois law mandates prescribers to

---

¶ According to federal law, a "legend" drug requires a prescription from a licensed practitioner and the following legend must appear on the manufacturer's label: "Caution: Federal law prohibits dispensing without prescription."

TABLE 1. Rates of DPS for Audited Prescriptions that were Legend, Brand Written and Formulary*

	Type of Pharmacy									
	Urban Community		Rural Community		Urban Chain		Rural Chain		Hospital Outpatient	
Pharmacist Behavior	N	%	N	%	N	%	N	%	N	%
Prescriptions drug product selected	145	15.3	27	6.33	43	11.78	21	9.10	10	5.00
Prescriptions dispensed as written	795	83.95	396	92.96	320	87.67	210	90.90	190	95.00

* Columns may not equal 100% because of rounding and because of instances of DPS outside the Formulary.

use a prescription blank which contains two signature lines and accompanying check boxes, one indicating "may substitute," the other indicating "may not substitute." Approximately 26 states have provision for the use of a two-line prescription format, but few, if any, besides Illinois, require the accompanying check boxes.** Because of this unique feature of the law, the effect of clear, unclear, confusing, and contradictory prescription signature formats was also examined.

Although the Illinois DPS law went into effect on July 1, 1978, the state allowed prescribers an additional, unspecified amount of time beyond that date to obtain prescription blanks containing the required 2-signature format. In the present study, when DPS occurred, only 36.6 per cent of the prescriptions examined were in format compliance on the effective date of the law (July 1, 1978). The rate of compliance rose, however, to 76.8 per cent by the end of the first year. Alternately, for those prescriptions that were dispensed as written, only 33.9 per cent were in compliance on the effective date of the law and this rose to 64.4 per cent after one year. Of all audited prescriptions, 55.25 per cent were found to be in compliance with the legal prescription format required. These data suggest that DPS occurred more frequently when the prescription format was in legal compliance.

Except in those cases where the prescriber has clearly indicated through signing and checking the accompanying box that DPS is allowed or prohibited, the intent of the prescriber may not be accurately communicated to the pharmacist under the Illinois law. If the prescriber signs either line but fails to check the boxes, signs one line but checks the alternate box, or continues to sign the bottom of the prescription as was done in the past, the pharmacist may be confused as to the prescriber's intent. Thus, the researchers created the labels of clear, unclear, confusing and contradictory to describe and further analyze the various signature formats encountered.

For all instances in which DPS occurred, the prescriber's instructions were in clear compliance only 9.9 per cent of the time. The remaining prescriptions were either unclear, confusing or contradictory in their instructions to the pharmacist, were telephoned prescriptions, or drug products were selected based upon the state's Public Aid program or on the basis of prior consent. Prescribers who either favor or oppose DPS could achieve their desired

---

** As of October 1980, 21 states and the District of Columbia required the physician to write a supplemental phrase on the prescription form such as "dispense as written—DAW," "No sub," "NS," "brand necessary" or "medically necessary" in order to prohibit drug product selection.

objectives by communicating more clearly to the pharmacist. The problem of lacking clarity is not a simple matter of prescriber intent but is more clearly a problem of the complex prescription format requirements that are on impediment to both the prescriber and the pharmacist.

Cost Savings

Of primary concern to consumers, legislators and health practitioners is the question of whether consumers saved money when drug product selection was used for prescriptions during the first year of the DPS law. Table 2 shows that significant savings occurred when DPS was practiced in Illinois. Further, it indicates that for private-pay prescriptions, an average of $1.30 was saved when drug product selection was used for a prescription, and that the state saved $1.88 on Public Aid prescriptions in the same way. The difference between the mean price for all three categories of the dispensed-as-written prescriptions and those where drug product selection was used (Table 2) shows that an average of $1.47 was saved every time DPS occurred. This average savings is consistent with savings reported from other states which range from $0.86 to $1.92 per prescription.[6,8] Again, notable differences in various state DPS laws prohibit an in-depth interpretation of these differences in savings in this paper.

A second methodology was employed to validate the reported savings. When the mean price for all products for which drug product selection was used is compared with the mean price for all products which were dispensed as written, an additional variable may contribute to the savings reported. If the drugs which were dispensed as written were largely more expensive drugs on the market, and if the prescrip-

TABLE 2. Analysis of Variance (ANOVA): Average Prescription Price by Payment Mechanism*

	Dispensing Behavior				
	Drug Product Selection (DPS)		Dispensed as Written (DAW)		
Payment Mechanism	Mean	N	Mean	N	Savings
Private pay	5.00	(1118)	6.30	(1750)	1.30
Public aid	4.60	(1220)	6.48	(281)	1.88
Other third party	8.14	(53)	6.72	(110)	−1.42
Total	4.87	(2401)	6.34	(2156)	1.47

Source	Degrees of Freedom	F Value
Dispensing Behavior (DB)	1	8.37†
DB at private pay	1	105.38†
DB at public aid	1	73.79†
DB at other third party	1	6.59†
Payment Mechanism (PM)	2	36.51†
PM at DPS	2	30.82†
PM at DAW	2	1.10
Dispensing Behavior by Payment Mechanism	2	25.17†
Within Cells	4526	

* Included audited and reviewed prescriptions which are legend, brand written and Formulary.

† $p \leq 0.01$

tions for which DPS occurred were generally for less expensive drugs, then the effect of this mix would confound the findings. To validate the reported savings of $1.47 per prescription, a modified matched-pair technique was used to ensure that the comparisons were based on the same drug entities.†† This technique showed that the savings for all entities for which matched-pairs were located was $1.45. This, compared to the finding of $1.47, served to validate the findings based on a comparison of mean prices for dispensed-as-written prescriptions versus those for which drug product selection was used.

The larger savings occurring with Public Aid prescriptions reflects two direct pricing influences upon the pharmacist. When a drug entity is single source, the Public Aid program reimburses the pharmacist an amount based on the Estimated Acquisition Cost (EAC) for that product. However, for those drugs that are multiple source, the pharmacist is reimbursed based on a Maximum Allowable Cost (MAC) established at the federal or state level which essentially reimburses for a lower priced generically equivalent drug, unless approval is obtained from the Department of Public Aid. In addition, the pharmacist may only dispense a generically equivalent product when so authorized by the prescriber, even though reimbursement for the drug product is fixed at the lower price. Such authorization occurs most often in the form of blanket authorization‡‡ from the prescriber for Public Aid patients. Thus, Public Aid policy in conjunction with a lowest-cost reimbursement approach virtually mandates that pharmacists use drug product selection for these prescriptions and assures the state of greater savings in drug costs. Complete reimbursement from the state for Public Aid prescriptions also includes a fixed dispensing fee per prescription dispensed.

Table 2 also shows analysis of variance (ANOVA) statistics for the reported mean prescription prices under two pharmacists' dispensing behaviors and three payment mechanisms. Main effects for dispensing behavior and payment mechanism, and the interaction of these two were found to be significant ($p \leq .01$). All three dispensing behaviors indicated by simple main effects that consumers saved significantly when DPS occurred.§§ Further, the Scheffe's method was employed to identify individual mean comparisons that were significant when DPS occurred. It was found that patients whose prescriptions were paid for by a third party were charged significantly ($p \leq .01$) more when DPS occurred as compared to private-pay and Public Aid patients. A systematic review of the third party prescriptions for which

---

†† For the last three months of the study period, when a prescription was legend, brand written and listed in the formulary, and when it was dispensed as written, the pharmacist was asked to price it as though drug product selection had been used. The pharmacist also was asked if he/she stocked the entity by another manufacturer, and if so, would he/she use drug product selection for this entity. The same procedure was carried out for those prescriptions for which drug product selection had been used and the pharmacist was then asked to price it as though it had been dispensed as written. To determine the reliability of this price estimate procedure, the pharmacist was first asked to furnish prices for 10 prescriptions that already were audited by the surveyor. These 10 estimated prices furnished by the pharmacist were compared via a paired t-test statistical method with the actual prices when the drug was dispensed. If the prices were determined to be reliable (t-test was not significant at the .05 level) the pricing procedure described above was completed. Only data from pharmacies with demonstrated reliability (consistency) in pricing were used for this matched-pairs study.

‡‡ Blanket authorization is a delegatory act by a prescriber in which prior consent is given to a pharmacist to practice DPS.

§§ When applied to the mean price for third party prescriptions that were dispensed as written versus drug product selected, the t-statistic for unequal variance was not significant. It was however, significant at $p \leq 0.01$ for both private pay and Public Aid prescription prices.

TABLE 3. Analysis of Variance (ANOVA): Drug Product Selection (DPS) Savings Occurring Based on Pharmacy Types*

	Urban Community		Rural Chain		Urban Chain		Hospital Outaptient		Rural Community	
Pharmacist Behavior	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
DPS Prescriptions	4.98	(1128)	4.08	(132)	4.35	(319)	5.08	(141)	5.75	(128)
Prescriptions Dispensed as Written (DAW)	6.80	(781)	5.84	(201)	5.60	(319)	6.00	(166)	6.32	(392)
Total Savings	1.82		1.77		1.22		.92		.62	

Source	Degrees of Freedom	F Value
Pharmacy Type (Type)	4	9.51‡
Type at DPS	4	6.84‡
Type at DAW	4	5.06‡
Dispensing Behavior (DB)	1	75.62‡
DB at urban community	1	133.01‡
DB at rural chain	1	20.88‡
DB at urban chain	1	20.94‡
DB at hospital	1	5.42‡
DB at rural community	1	2.63‡
Pharmacy Type by Dispensing Behavior	4	2.73†
Within Cells	3798	

* Includes audited and reviewed prescriptions which are legend, brand written and Formulary.

† $p \leq 0.05$

‡ $p \leq 0.01$

product selection was used could not fully explain this phenomenon. Various copayment and deductible features have been shown to influence both prescribing and dispensing behaviors when a third party carrier is involved, but data to confirm such an influence were not collected.

In examining additional variables that contribute to consumer drug costs, pharmacy type and pharmacy location were each identified for further analysis. It was anticipated that these variables might help to explain price variations to consumers and public policy makers. Using a 2 × 5 ANOVA procedure comparing prices by types of pharmacies, significance of the overall savings was established ($p \leq .01$) and significance between types of pharmacies was shown ($p \leq .01$). Simple main effects were used to show the significance of savings within the five types of pharmacies. When DPS occurred, and the savings were calculated for all five cells, urban community pharmacies showed the highest savings ($1.82 per prescription), ($p \leq .01$), and rural community pharmacies showed the lowest savings ($0.62). Table 3 shows the calculated savings for each of the five types of pharmacies studied.

To further contrast urban versus rural and community versus chain pharmacy prescription prices, additional Scheffe's tests were calculated.[|||] The results showed that when DPS did not occur, prescriptions

---

[|||] The mean prices used for these comparisons are not shown in Table 3. They represent data generated by combining community and chain price data in both urban and rural environments, and by combining community price data and chain price data from both environments.

dispensed in urban pharmacies were not significantly more expensive than prescriptions dispensed in rural pharmacies ($6.45 vs $6.20). Also, when DPS did not occur, prescriptions dispensed in community pharmacies were significantly more expensive on the average, than those dispensed in chain pharmacies ($6.66 vs $5.70), ($p \leq 0.05$). Similarly, when DPS did occur, the differences in prices between urban and rural pharmacies were not significantly different ($4.85 vs $4.91). Finally, when DPS did occur, prescriptions dispensed in community pharmacies were significantly more expensive than those dispensed in chain pharmacies ($5.06 vs $4.27) ($p \leq .05$). A meaningful comparison of hospital outpatient prescription prices could not be made because of the lower purchase prices hospital pharmacies may enjoy for bona fide outpatient dispensing purposes under the Nonprofit Corporations amendment to the Robinson-Patman Act and a 1976 Supreme Court interpretation.

In summary, whether DPS occurred or not, there was no significant difference in prescription prices between urban and rural pharmacies when chain and community prescription prices were combined. Alternately, whether DPS occurred or not, there was a significant difference in the average prescription prices between those dispensed in community versus chain pharmacies when urban and rural prescription prices were combined. Thus, consumers can expect to pay a variable amount for their prescriptions depending upon the type of pharmacy patronized, but the location of their pharmacy in either an urban or rural area, was not shown to affect the price of their prescriptions significantly.

Conclusions

An evaluation of Illinois' DPS law demonstrated that it enabled consumers to save, but that various factors still impede acceptance of this new role by pharmacists and physicians. It is evident that the restrictive nature of the positive formulary (listing 172 drug entities) and the prescription format requirement of the Illinois law combine to decrease the potential for DPS by pharmacists.¶¶ This legislated change in pharmacist and physician behavior is affected by a multitude of additional attitudinal and perceptual factors, some of which were analyzed as part of this evaluation study, but because of length will be reported in additional publications.

The data presented from this research show that the rate of DPS varies between types and locations of pharmacies and is affected by the mix of private pay, Public Aid and third party prescriptions dispensed. In addition, both prescriber compliance with the mandated prescription format and the clarity of instructions to the pharmacist appeared to affect the DPS behavior of the pharmacist.

Consumer cost savings on prescription drugs are occurring in Illinois and these savings appear to vary, depending upon the source of payment for the prescription. Consumers patronizing urban community pharmacies in Illinois can expect to pay more for their prescriptions, but also can expect to save more when DPS occurs. Finally, it was concluded that prescription prices were affected by the type of pharmacy patronized but were not affected by the urban or rural location of the pharmacy.

There is a continued need to conduct further research in this area to determine those features of the various DPS laws which afford consumers access to less expensive prescription drugs, while maintaining drug quality and other professional prerogatives of both prescribers and pharmacists. The difficulty of meeting the needs of consumers, health professionals and the drug manufacturers is recognized, but not always addressed in the legislative process. Hopefully, continued research in the area will facilitate problem solving in the public interest.

¶¶ An amendment to the Illinois DPS law which simplifies the prescription format by requiring one signature line accompanied by two check boxes designated "may substitute" and "may not substitute" has been enacted recently. The present study provided much of the data utilized in promoting this amendment.

References

1. Federal Trade Commission. Drug product selection. Staff report, Bureau of Consumer Protection. Washington, D.C.: Federal Trade Commission, 1979.

2. Reich RB. Drugs and medical devices policy session briefing book. Washington, D.C.: Federal Trade Commission, 1978.

3. Anonymous. Drug substitution: where does your state stand. Nat Assoc Retail Druggist J 1980;102:44.

4. Francetic I, Lasagna L, Weintraub M, Karch F. Prescription prices under the New York generic substitution law. Ann Intern Med 1980;92:419.

5. Goldberg T, Moore WE, Koontz T, et al. Evaluation of impact of drug substitution legislation. J Am Pharm Assoc 1976;NS16:66.

6. Goldberg T, Aldridge GW, DeVito CA, Vidis J, Moore W, Dickson WM. Impact of drug substitution legislation: a report of the first year's experience. J Am Pharm Assoc 1977;NS17:216.

7. Goldberg T, DeVito CA, Smith D, et al. Evaluation of economic effects of drug product selection legislation. Med Care 1979;17:411.

8. Vuturo GJ, Krischer JP, McCormick WC. Drug product selection: the Florida experience. Am J Public Health 1980;70:479.

9. Goldberg T. Comparison of key findings relevant to the effectiveness of generic drug substitution laws in four states. Presented to the Economic, Social and Administrative Sciences Section of the Academy of Pharmaceutical Sciences, American Pharmaceutical Association Annual Meeting, Washington, D.C., May, 1980.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Podiatric Surgery: The Need for a Second Opinion

MADELON LUBIN FINKEL, PH.D.,* EUGENE G. MCCARTHY, M.D., M.P.H.,† AND DAN MILLER‡

A second-opinion podiatry program, designed to contain costs and enhance the quality of care, was established in the greater New York City area by the Building Service Employees' Health Fund. All nonemergency foot surgery recommended by a podiatrist that cost over $100 was included in this program. Of the 363 individuals who had a second opinion, 82 per cent were confirmed for surgery; however, the confirmation rate among those seeing a podiatrist serving as consultant was 94.3 per cent, while the confirmation rate of those seeing an orthopedist serving as consultant was 49.5 per cent. For every condition, the podiatrists overwhelmingly confirmed the need for surgery. Conversely, the orthopedists were more conservative in their recommendations regarding surgery. The program's cost was $311,813. If a second opinion consultation wasn't required for removal of corns or ingrown toenails, the total would be much less. It seems sensible (economically and medically) to encourage or require a consultation with an orthopedist for conditions such as bunions, hammertoes and heel spurs, and not to require a second opinion consultation for superficial podiatric conditions such as corns and ingrown toenails, which are treated predominantly by the podiatrist.

THE FOOT is probably the most used and the most neglected part of the human body. Each day an average person's feet hit the ground 20,000 times with a force that is three times one's body weight. While most people begin life with two normal feet, problems can and often do, develop from abuse or neglect. Many of these problems can be prevented with proper foot care. Yet, many regard the foot as being an undramatic part of the body and regard its problems as trivial. Many prefer not to seek proper medical or podiatric attention; however, such neglect may lead to disabling conditions. Not surprisingly, the current popularity of jogging and marathon running has created a surge of complaints concerning foot, ankle and knee problems.

At least 10 million individuals suffer from soft tissue complaints (corns, calluses, bunions and nail disorders) and/or static foot deformities (clubfoot, flatfoot). Of all the things that can go wrong with the 26 bones, 107 ligaments and 19 muscles that make up the foot, none is more common than the development of corns. Corns, simply a small callus with a compact central core, result from pressure of the hard mass against underlying tissue. The best way to get rid of corns is to eliminate the original pressure or friction that caused

---

* Assistant Professor, Department of Public Health, Cornell University Medical College, New York, New York.

† Clinical Professor, Department of Public Health, Cornell University Medical College, New York, New York.

‡ Public Health Intern (formerly) and Master's Candidate, Columbia University School of Public Health, New York, New York.

Preliminary Findings were presented at the Annual Meeting of the American Federation of Clinical Research, San Francisco, California, April 1981.

Address for reprints: Madelon Lubin Finkel, Ph.D., Department of Public Health, Cornell University Medical College, 1300 York Avenue, New York, NY 10021.

0025-7079/82/0800/0862/$01.25 © J. B. Lippincott Co.

them; i.e., get better fitting shoes. Bunions, resulting from inflammation of the joint at the base of the big toe, are one of the most common afflictions of the normal adult foot. They are four times more frequent in women than in men, primarily because women's shoes increase the valgus strain on the big toe, thus irritating the condition. The underlying cause of most bunions is improper alignment of the bones of the foot and irritation by the shoe serves to aggravate the condition. Many spend thousands of dollars having a bunion corrected, only to have it recur because the faulty bone alignment remains. Ingrown toenails, too, are a major source of foot problems. Improper cutting of the nail or improper footwear can cause this condition. It is estimated that flatfoot and soft tissue complaints are not adequately cared for, either because the individual patient does not seek care or because inadequate attention is paid to these conditions once care is sought. Conversely, most static foot deformities receive professional attention.[1]

A survey of foot problems in the continental U.S. found that 74 million Americans (40 per cent of the population) reported they have foot problems.[2] Those most prone to foot trouble are women, the elderly, those with incomes below $5000, nonwhites and those residing in the South. Roughly 23 million have had surgery performed on one or both feet and 13.5 million have had some form of medical treatment for their conditions. Simple arithmetic shows that almost 38 million people wih foot problems have had neither surgery nor medical treatment.

Most surgery on the foot is elective (nonemergency) and is performed to correct deformity, relieve pain, improve function or prevent future disability. The increase in the rate of specific podiatric surgical procedures clearly reflects the difficulty people are having with their feet. From 1971 to 1978, there was a 558.8 per cent increase in the rate of foot surgery in which the bones of the foot were cut. There was a 103.9 per cent increase in the rate of joint surgery performed on the foot or toe during this same period. In comparison, the rate of all orthopedic surgery increased 36.3 per cent from 1971 to 1978 and the rate of all operations increased 24.3 per cent during this same time period.[3]

Foot problems are handled by either an orthopedic surgeon or a podiatrist. While the former is a surgeon who specializes in treating the musculoskeletal system of the body, including the foot, the podiatrist is concerned with the diagnosis and treatment of the foot and ankle. Orthopedists see a greater percentage of foot problems related to injury, as well as static foot deformity such as clubfoot, while podiatrists treat the overwhelming majority of soft tissue complaints and flatfoot. Disorders of the toenail, for example, are probably diagnosed and treated more often by podiatrists than by any other medical specialist. In most states, podiatrists are allowed by law to deal only with the foot; however, in six states they are permitted to treat up to the knee, and in another six states they may work on the foot and tendons of the leg.[4] Most podiatrists still confine themselves to work on the foot, however. The situation is analogous to that of a nonsurgical physician who technically is licensed to set a broken bone, but he is unlikely to do so.

Podiatrists have considerable patient acceptance in the United States. Roughly 7500 podiatrists engaged in patient care of some 35 million visits in 1978, a 40 per cent increase since 1970. Although the demand is increasing, the supply of podiatrists is relatively small. While there is one medical doctor for every 670 people and one dentist for every 2,000 individuals, there is only one podiatrist for every 25,000 persons.[5]

Given the extent of foot problems among the population, it is not surprising that the cost of foot care in the United States is staggering. It has been estimated that the provision of foot care (excluding home treatment using over-the-counter remedies) costs over $1 billion a year with office visits and hospital stays accounting for $450 million each. The remainder is accounted for by surgeons' fees, emergency room care and outpatient clinics. Foot-related office visits account for about two per cent of all payments to direct providers of health care, and foot-related hospital stays account for roughly one per cent of all hospital expenditures.[1] Ironically, cost containment programs have ignored podiatric services. One innovative program, designed to contain costs related to foot care, was established in the greater New York City area by the Building Service Employees' Health Fund, Local 32B-J.

Study Population

The Service Employees International Union Local 32B-J represents workers of both sexes who are employed in the maintenance, operation and upkeep of all private and public buildings (residential and commercial) in boroughs of the City of New York, except the Bronx. Those employed as superintendents working in loft and office buildings, exterminators and window cleaners are excluded specifically. The local is chartered by the Service Employees International Union under the American Federation of Labor-Congress of Industrial Organizations—600,000 members and dependents. It is composed of 120,000 members and dependents and serves as the exclusive bargaining representative for its members, assisting them to obtain adequate compensation for their labor and promoting their welfare.

Members work as elevator operators, doormen, porters and office cleaners in over 6,000 scattered worksites. Sixty per cent work the night shifts, particularly the women who work as cleaning ladies, usually without supervision. Since the tasks of maintaining and servicing buildings require that the individual to be on their feet much of the time, it is not surprising that there is a high incidence of foot problems among the membership.

Mandatory Second Opinion Podiatry Program

From 1979 through 1980, the Building Service Local 32B-J Health Fund paid over $288,000 for podiatric surgery. Almost $251,036 was paid for in-hospital surgery and $37,727 was paid for in-office surgery. The majority of in-hospital claims were for bunion removal while the majority of the in-office surgery was for removal of ingrown toenails. It was decided to require a second opinion consultation for podiatric conditions in order to possibly contain the costs of such surgery and to hopefully enhance the quality of care.

The second-opinion program for podiatric problems was established in the fall of 1979. Each member and his or her dependents is required to schedule a second opinion consultation for all nonemergency foot surgery that is recommended by a podiatrist costing over $100. Whether the surgery is to be performed on an inpatient or outpatient basis is immaterial. The consultant (podiatrist or orthopedist) is paid $50 for the consultation and the Fund pays the costs of any ancillary tests performed by the consultant. Although the program tried to refer half of the individuals to a podiatrist and half to an orthopedist for the second-opinion consultation, it became quite obvious that most orthopedic surgeons refuse to see patients with common podiatric problems such as ingrown toenails, corns, warts and other minor conditions.

It was the Health Fund's policy not to indicate to the second-opinion consultant (podiatrist or orthopedist) the initial diagnosing physician's specialty. The initial diagnosing physician's name is listed on an evaluation form sent to the consultant, but the degree (MD, DO, DPM) is not, therefore, unless the second-opinion consultant happens to know that Dr. X is a podiatrist, the consultant would not be able to tell whether a podiatrist or an orthopedist initially recommended surgery. It is possible that the consultant could have asked the patient directly whether a podiatrist or an orthopedist recommended surgery. Each consultant placed on the second-opinion consultant panel, however, has agreed to provide an unbiased evaluation of the problem and we have no reason to expect that a consultant purposively would reject the opinion of a "competitior."

Findings

Demographic Profile

Of the 363 individuals who had a second opinion for initially recommended foot surgery, 61.4 per cent were females. The age of the sample ranged from 12 years (a covered dependent) to 70 years with a mean age of 47.6 years. There were no statistically significant differences in age between the sexes.

Among the males in the study, 22.9 per cent were porters, 16.4 per cent were doormen, and an equal percentage were handymen, elevator operators or individuals for whom occupational status was unknown, (12.1 respectively). Among the females in the study, 48.9 per cent were cleaning ladies and 23.3 per cent were covered dependents, the overwhelming majority of whom were nonworking housewives. The remaining individuals of both sexes were either building superintendents or building security guards.

One third (30 per cent) of the sample initially were recommended for a bunionectomy (bilateral or unilateral), and 22 per cent initially were recommended for surgical correction of hammertoes, two common orthopedic conditions. Roughly 15 per cent initially were recommended for surgical removal of an ingrown toenail. The remainder were initially recommended for other common podiatric conditions.* Of those initially recommended for common orthopedic conditions, over half (59.9 per cent) had multiple diagnoses; i.e., removal of bunion and correction of hammertoes. Of those initially recommended for common podiatric conditions, 21.2 percent had multiple diagnoses, i.e., removal of corns and warts.

Second-Opinion Consultant's Recommendation

Approximately one-quarter (27.8 per cent) of those initially recommended for podiatric surgery were seen by an orthopedic surgeon (Board Certified) and 72.2 per cent were seen by a podiatrist. Roughly 82 per cent of the study population was confirmed for surgery. There were no statistically significant differences in the confirmation rate between the males and the females or among age groups. An almost identical percentage within each age and sex category were confirmed for surgery.

A closer inspection of the data shows that the confirmation rate among those seeing a podiatrist was 94.3 per cent as compared to a confirmation rate of 49.5 per cent among those who were seen by an orthopedist. Whereas the podiatrists did not confirm the need for surgery in 5.7 per cent of the cases, the orthopedists did not confirm the need for surgery in 50.1 per cent of the cases.

Data presented in Table 1 show that, with the exception of those initially recommended for surgery to correct a heel spur, the overwhelming majority within each podiatric diagnostic category were confirmed for surgery.

When the data were analyzed by the second opinion consultant's specialty, dramatic differences were apparent. Table 2 shows the confirmed/not confirmed status

---

* Corns, metatarsalgia, warts, neuroma.

TABLE 1. Podiatric Conditions Analyzed by Confirmed/Not Confirmed Status

Condition	Confirmed		Not Confirmed		Total	
	N	%	N	%	N	%
Bunion	88	80.0	22	20.0	110	100.0
Corn	24	92.3	2	7.7	26	100.0
Hammertoe	75	86.2	12	13.8	87	100.0
Heel spur	10	43.5	13	56.5	23	100.0
Ingrown toenail	54	100.0	0	0	54	100.0
Neuroma	10	90.9	1	9.1	11	100.0
Osteoarthritis	6	75.0	2	25.0	8	100.0
Other	31	70.1	13	29.5	44	100.0
Total	298	82.1	65	17.9	363	100.0

of the podiatric conditions screened by the podiatrists. For every condition, the podiatrists overwhelmingly confirmed the need for surgery. Conversely, Table 3 shows that the orthopedic second-opinion consultant was less likely to confirm the need for surgery in each category. Whereas the podiatrists confirmed the need for a bunionectomy in 95.8 per cent of the cases, the orthopedists confirmed 50 per cent of the potential bunionectomies. Whereas the podiatrists confirmed the need for surgery in 93.7 per cent of the cases with hammertoes, the orthopedists confirmation rate for hammertoes was 66.7 per cent. Whereas the podiatrists confirmed the need for repair of a heel spur in 75.0 per cent of the cases, only one individual was confirmed for this operation by an orthopedic consultant of the eight cases seen.

Table 4 shows the distribution of consultant reasons for nonconfirmation analyzed by podiatric conditions. In one third of the the cases, the consultants felt that corrective footwear was preferable to surgery and in almost 17 per cent of the cases, the consultants felt that the particular procedure recommended initially was incorrect. Of those not confirmed for a potential bunionectomy, in 27.8 per cent of the cases the second-opinion consultants felt that the initial procedure recommended was incorrect and in 22.2 per cent of the cases corrective footwear was recommended instead of surgery. In 38.5 per cent of the cases, of those not confirmed for

TABLE 2. Confirmed/Not Confirmed Status Analyzed by Podiatric Conditions Screened by a *Podiatrist*

Condition	Confirmed		Not Confirmed		Total	
	N	%	N	%	N	%
Bunion	69	95.8	3	4.2	72	100.0
Corn	17	94.4	1	5.6	18	100.0
Hammertoe	59	93.7	4	6.3	63	100.0
Heel spur	9	75.0	3	25.0	12	100.0
Ingrown toenail	52	100.0	0	0	52	100.0
Neuroma	10	100.0	0	0	10	100.0
Osteoarthritis	6	100.0	0	0	6	100.0
Other	26	89.7	3	10.3	29	100.0
Total	248	94.7	14	5.3	262	100.0

TABLE 3. Confirmed/Not Confirmed Status Analyzed by Podiatric Conditions Screened by an *Orthopedist*

Condition	Confirmed		Not Confirmed		Total	
	N	%	N	%	N	%
Bunion	19	50.0	19	50.0	38	100.0
Corn	7	87.5	1	12.5	8	100.0
Hammertoe	16	66.7	8	33.3	24	100.0
Heel spur	1	9.1	10	90.9	11	100.0
Ingrown toenail	2	100.0	0	0	2	100.0
Neuroma	0	0	1	100.0	1	100.0
Osteoarthritis	0	0	2	100.0	2	100.0
Other	5	33.3	10	66.7	15	100.0
Total	50	49.5	51	50.5	101	100.0

correction of a heel spur, the consultants felt that corrective footwear was preferable. Reasons for nonconfirmation varied more among those recommended initially for correction of hammertoes, but the data should be interpreted with caution due to the small numbers involved.

Differences were apparent in reasons for nonconfirmation by the second-opinion consultant between the orthopedists and the podiatrists. In general, the orthopedists were more conservative in their recommendation for surgery. In 19 out of 20 cases, it was the orthopedist who suggested that corrective footwear was preferable to surgery. In 9 out of 11 cases, the orthopedist felt that the initially recommended procedure was incorrect and that the specifically recommended procedure should not be performed. In four out of six cases, the orthopedist felt the surgical procedure was contraindicated, and in five out of six cases, the orthopedist felt that the symptoms were not severe enough to warrant surgery. As was shown earlier, the podiatrists were more likely to confirm the need for surgery.

The decision to have surgery is made by the patient, regardless of the consultant's recommendation. Of those who were not confirmed, 81.5 per cent did not have surgery. Conversely, of those who were confirmed for surgery, the majority (60.1 per cent) did not have the operation performed. There was no statistical difference in surgical status between the males and females. Data show that 52.9 per cent of those seen by a podiatrist decided to have surgery, whereas 66.3 per cent of those seen by an orthopedist decided against surgery. The majority of those who did have the surgery performed did so within two months from the date of their consultation (76.4 per cent), and 90 per cent had the surgery performed within three months. In 95 per cent of the cases in which surgery was performed, the surgeon was a podiatrist. Of those who had surgery, 52.9 per cent had the procedure performed in an office. Conditions such as corns, ingrown toenails, removal of warts and other minor podiatric conditions were performed on an outpatient rather than on an inpatient basis.

In general, the podiatrists ordered more ancillary tests than did the orthopedists. Podiatrists ordered additional x-rays in 38.3 per cent of the cases, compared to the orthopedists who ordered additional x-rays in 11.5 per cent of the cases. Each patient was instructed to bring x-rays with them to the second-opinion consultation. Since most of the patients were seen initially by one of the Fund's Health Center doctors, we can state with assurity that the x-rays were handed to the patient a day before the second-opinion consultation. Others had to obtain their x-rays from their own private physicians or from the hospital. Each patient is reminded to bring x-rays to the consultation during a telephone conversation by one of the second-opinion referral officers. However, of the 72 potential bunionectomies seen by the podiatrists, 29 (40.4 per cent) had additional x-rays taken. Conversely, of the 38 cases of potential bunionectomies seen by the orthopedists, two (5.3 per cent) needed additional x-rays.

Of the 63 individuals with hammertoes seen by the podiatrists, 32 (50.8 per cent) had additional x-rays taken. Conversely, only two of the 24 individuals with hammertoes seen by the orthopedists needed additional x-rays taken (8.3 per cent). Of the 12 individuals with a heel spur seen by the podiatrists, (N = 8; 66.7 per cent) had additional x-rays taken as compared to two of the 11 individuals with a heel spur seen by the orthopedists (18.2 per cent).

Of 18 individuals with corns seen by the podiatrists, five (27.8 per cent) had x-rays taken as part of the second-opinion examination. None of the eight individuals with corns who were seen by an orthopedists needed additional x-rays. The cost of the additional x-rays taken by the podiatrists total $3,700.00.

Economic Considerations

The surgical bill for the major podiatric conditions performed in hospitals§ totaled $104,071; the mean cost was $1169.00. For those procedures performed in the office (outpatient) the surgeons' bills totaled $31,717; the mean was $488. The average total expense for inhospital podiatric surgery totaled $2,921 while the average total expense for in-office surgery totaled

TABLE 4. Reasons for Nonconfirmation Analyzed by Podiatric Conditions

Condition	Corrective Footwear Preferable		Incorrect Procedure		Symptoms Not Severe Enough		Surgery Contraindicated		Patient Asymptomatic		Medication Preferable		Other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Bunion	4	22.2	5	27.8	2	11.1	3	16.7	2	11.1	0	0	2	11.1	18	100.0
Corn	1	20.0	0	0	2	40.0	0	0	0	0	2	40.0	0	0	5	100.0
Hammertoe	1	16.7	2	33.3	0	0	1	16.7	1	16.7	0	0	1	16.7	6	100.1
Heel spur	5	38.5	0	0	1	7.7	2	15.4	0	0	1	7.7	4	30.8	13	100.1
Neuroma	0	0	1	50.0	0	0	0	0	0	0	1	50.0	0	0	2	100.0
Osteoarthritis	0	0	1	33.3	0	0	0	0	1	33.3	0	0	1	33.3	3	99.9
Other	9	50.0	2	11.1	1	5.6	0	0	0	0	0	0	6	33.3	18	100.0
Total	20	30.8	11	16.9	6	9.2	6	9.2	4	6.2	4	6.2	14	21.5	65	100.0

---

§ Bunions, removal of corns, hammertoes, heel spur.

TABLE 5. Surgical Bills for Specific Podiatric Conditions

Condition	Bill	Average Bill	Range Hi	Range Lo
Bunion	$ 90199.00	$1610.70	$3091.00	$300.00
Corn	5472.00	390.86	967.00	280.00
Hammertoe	33345.00	741.00	2352.00	165.00
Heel spur	4547.00	757.83	1285.00	410.00
Ingrown toenail	5775.00	160.42	390.00	55.00
Other*	22748.00	733.81	2313.00	125.00
Total	$162086.00	$ 862.16		

* Neuroma, osteoarthritis, other, not specified.

$494. Table 5 shows the costs for the major podiatric conditions. The surgeons' bills for bunionectomies totaled $90,199, with a mean of $1,610.70 and a range of $300 to $3,091. The bills for removing corns averaged $390.86 and for removing of ingrown toenails averaged $160.42. The surgeons' bills for correction of hammertoes totaled $33,345, with a mean of $741.00 and a range of $165 to $2,352. the surgeons' bills to correct heal spurs totaled $4,547.00 with a mean of $757.83 and a range of $410 to $1,285.

One explanation for the large range in bills for specific procedures is that a bilateral bunionectomy was performed or that several toes were corrected at the same time. Also, several procedures were performed at one time but were counted as one procedure (bunionectomy, osteotomy, tendon lengthening and capsulotomy performed at one time). Some individuals also suffered complications, which served to inflate the surgeon's bill.

Since the Fund pays "reasonable and customary" costs, it is likely that the patient would have to pay the difference between what is charged by the surgeon and what is paid for or reimbursed by the Fund. We found that the podiatrists, in particular, were charging in excess of the reasonable and customary costs, but were more than likely to accept the amount paid by the Fund. Specifically, the podiatrists billed a total of $64,870 for in-office surgery, but accepted payments totaling $36,652. For example, the Fund was billed $8,415 for ingrown toenails, but only paid $5,635. The podiatrists billed a total of $240,755 for in-hospital surgery, but accepted $104,071—less than half of what they billed! The average bill was $2,700 and the average reasonable and customary payment was $1,169.

The cost of the consultations totaled $18,100, the reimbursement of the surgeries performed[||] totaled $288,763 and reimbursement of ancillary tests totaled $4,950, for a grand total of $311,813. If a second-opinion consultation is not required for removal of corns, warts and ingrown toenails, the cost of the program would be much less.

Discussion and Conclusion

Until fairly recently, most podiatrists could not obtain hospital privileges and were content to treat the outer tissues. Now, podiatrists are pressing to operate in hospitals and to treat the whole foot (muscles, bones, nerves, tendons and ligaments). Financial motivation certainly plays a part, as does the upgrading of the

---

[||] In and out of hospital.

instruction in podiatry schools. Since major third-party coverage does not reimburse for routine podiatry care, surgery has become the focus of podiatry's push.[6] Operating on the foot brings in more money than trimming and paring. Since most states will not permit podiatrists to administer general anesthesia to office patients, podiatrists must rely on local anesthesia, using minimum incision surgery.

Orthopedists contend that such surgery can cause substantial trauma to surrounding tissues and that the bone being operated upon is not visible. Naturally, they feel that the orthopedist, not the podiatrist, is the most competent to operate on the foot. Orthopedists do have more extensive surgical and medical training and are better able to handle complications.

Even though many foot problems do not require surgery, foot surgery is a very lucrative business. Statistics show that millions of individuals have had their foot problems treated medically and/or surgically, at a cost of millions of dollars, to relieve foot pain or discomfort. The findings from the 32B-J podiatry second-opinion program clearly show the predilection among podiatrists to confirm podiatric surgery and to order an excessive amount of ancillary tests prior to and after surgery. Analysis of the program's first year has made it crystal clear that thousands of dollars are being spent on questionable podiatric surgery. Since the orthopedic consultants refused to examine ingrown toenails and corns, and since almost all second-opinion podiatrists confirmed the need for surgery for these procedures, the Fund decided in April, 1981, to no longer require a second opinion for these procedures.

Regarding the more complicated podiatric conditions (the bunionectomies, the hammertoes and the heel spurs), a second-opinion consultation definitely should be required. It is our opinion that the consultant should be an orthopedic surgeon. These procedures usually are performed in the hospital and have high rates of nonconfirmation. They also are the more costly procedures and substantial savings could be realized if a second-opinion consultation was required. It is also clear that the podiatrists in this study tended to take advantage of a situation and not only were charging excessive fees, but probably were performing many procedures unnecessarily. In April 1981, the Fund also announced that podiatrists were no longer eligible to serve as second-opinion consultants. Only orthopedists are utilized as consultants for all foot-related problems for which a second-opinion consultation is required.

References

1. Greenberg L. An assessment of foot health problems and related health manpower utilization and requirements. J Am Podiatry Assoc 1977;67:102.
2. Gould N, Schneider W, Ashikaga T. Epidemiological survey of foot problems in the continental United States: 1978–1979. Foot and Ankle 1980;1:8.
3. National Center for Health Statistics, Hospital Discharge Survey, Personal communication.
4. Mattera MD. Podiatrists: Pushing for equal rights on the medical team. Medical Economics 1979 (Oct. 15), 159.
5. Fowler EM. Demand likely for podiatrists. New York Times 1980 (Sept. 10), D 19.
6. What to do when your feet hurt. Consumer Reports, March 1981, pp 142-146.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Utilization of an Automated Multiphasic Health Testing System for Performing Prehospitalization Examinations

RAFAEL S. CAREL, M.D., M.P.H.,* ERNESTO KAHAN, M.D., M.P.H., YAAKOV HART, M.D., AND NAOMI PANUSH, M.SC.

Various impacts of performing prehospitalization examinations (of elective surgical patients) by utilizing automated multiphasic screening technology were evaluated. Compared to the conventional way of conducting such examinations (in the concerned health care organization) the automated system was found to be more efficient (results were available earlier, in a more complete and convenient format), saved considerable time to patients and resulted in a shorter average hospital (postoperative) stay, suggesting that the automated technology is also cost effective. Postoperative stay was reduced by 20 per cent (1.8 days) in the group of patients who undertook the automated multiphasic testing, as compared to a similar group of patients who performed the required tests in the conventional way. Possible explanations for this finding are discussed.

DIFFERENT HEALTH CARE organizations have approached and solved the problem of elective prehospitalization testing in different ways. The health care organization under consideration is a large, prepaid, comprehensive health plan which has been described previously.[1] To facilitate and shorten hospital stay prior to the planned operation, each surgical ward has established a routine battery of tests as a prerequisite for admission. This battery includes the following: physical examination; blood pressure; ECG; chest roentgenogram; bleeding, clotting and prothrombin time; basic blood tests (CBC, urea, glucose, electrolytes, etc.) and urinalysis. Specific tests related to evaluation of the major illness such as x-ray series, biopsies or cytological examinations usually are performed in outpatient (specialist) clinics. The routine, basic prehospitalization examinations are conducted under the supervision of the family physician in the regular clinics and laboratories of the concerned health organization. Hospitalization and surgery dates are scheduled well in advance (by the specific surgical service) for most patients. Each ward is allocated fixed days for using operating theaters and scheduling of elective cases is planned accordingly.

This study was to evaluate the currently used method of executing prehospitalization examinations and to compare it with

---

From the Department of Preventive and Social Medicine, Sackler School of Medicine, Tel-Aviv University, Israel.

* Dr. Carel is also from the MOR Institute for Medical Data.

Reprint requests: R. S. Carel, M.D., Department of Preventive and Social Medicine, Sackler School of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel.

0025-7079/82/0800/0871/$01.05 © J. B. Lippincott Co.

performing the same battery of tests by utilizing an automated multiphasic health testing system (AMHTS).

Material and Methods

Figure 1 is a block diagram of the various steps necessary for completing the required tests by both modalities studied—the conventional one (steps 1 to 3) and the automated one (steps 1 and 2). In the conventional "track," the family physician refers the patient to several different laboratories (step 2) to acquire all the necessary medical data included in the prehospitalization profile, while in the AMHTS track all these tests are performed in one facility (step 1) by a dedicated unit.

The study group included 155 subjects. Seventy-eight patients (63 males and 15 females) were tested in the conventional way and 77 (63 males and 14 females) by the AMHTS. Each study group included all cases scheduled for elective surgery in one urological ward during six consecutive months (3 months for each group). The concerned ward draws its patients from a defined geographic area and a uniform mixture of cases is scheduled through the whole year. Age distribution of the subjects is presented in Table 1, showing a similar distribution for both groups. Mean age in the conventional group was 66.1 years and that of the AMHTS group was 63.0 years. Most of the subjects in both groups were over 60 years old (about 70 per cent). Distribution of examinees by types of diagnoses (ICDA definitions) is given in Table 2. About 40 per cent of the subjects in both groups were admitted for prostatectomy due to benign hypertrophy of the prostate and another 10 per cent for malignant tumor of the prostate. Approximately 20 per cent were hospitalized because of other suspected genitourinary malignancies. All other surgical problems constituted less than 30 per cent for both groups. Thus subjects in both groups were quite similar with respect to age and sex distribution, types of diagnoses, severity of illness, residential area and socioeconomic status, as verified by a questionnaire, indicating that selection bias could not be a major factor.

Results

Patients who were examined by the conventional method required, on the average, two work days to complete the necessary tests. Thirty people (39 per cent) completed the examinations in 1 day, 19 (24 per cent) in 2 days and 17 (22 per cent) in 3 days. In 12 subjects, exact information could not be obtained. All the examinees who were tested in the AMHTS completed

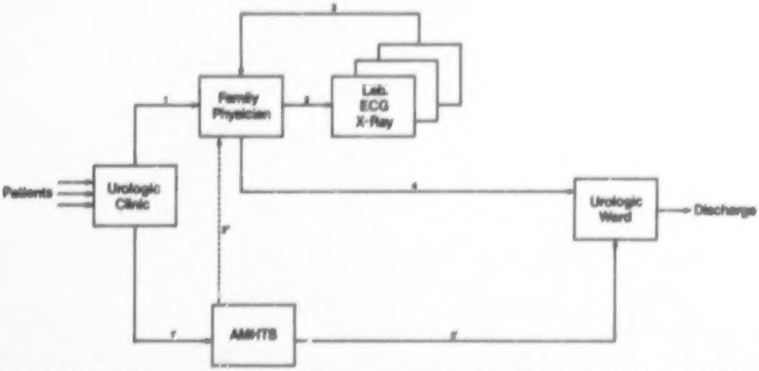


FIG. 1. Flow chart of the steps involved in conducting preadmission examinations by the conventional and AMHTS methods.

the procedure in less than 4 hours (34 in less than 2 hours, 36 in less than 3 hours and 4 in 3–5 hours). The major reason for the long duration of the conventional examinations was that patients had to apply to different laboratories, which often were in different locations. For performing the various tests, 40 patients (51 per cent) visited only one facility, 18 (23 per cent) visisted two facilities and 20 (26 per cent) visited three facilities. For this group, final results of the various tests were received by the family physician (step 3, Fig. 1) within 2 weeks with 74 (94 per cent) patients, 2–3 weeks with two (3 per cent) patients and with another two in over three weeks. On the average it took 8.1 days until the family physician had all the required test results at this disposal. In the AMHTS method all results were collected, collated and printed as a uniform report by a computer within 3 days and transferred directly to the concerned urological ward. It should be realized that hospitalization (and operating) dates were not affected in any way by the method of performing preadmission examinations. Each patient had a fixed admission date independent of how soon the required tests were obtained (or the method of prehospitalization testing). Therefore, even though all patients tested by the AMHTS had their results ready within 72 hours, only 45 (58 per cent) were hospitalized within 2 weeks of their prehospitalization examinations.

No difference between the two groups, with respect to the time interval between admission and operation, could be demonstrated. This could be expected because elective admission date and the day of surgery are predetermined for each surgical service in the concerned hospital (and usually are independent of the prehospitalization tests). Almost all of the examinees in both groups were admitted on their scheduled date.

In Figure 2, the cumulative distribution of the time interval between day of surgery and day of discharge is shown. Fifty per cent of AMHTS subjects were discharged less than 7 days after their operation, while in the other group only 30 per cent of the patients were discharged 7 days after surgery. The average postoperative stay in the AMHTS group was 6.9 days and that in the "conventional" group was 8.7 days (1.8 days longer). The difference between the means of postoperative stay of both groups is statistically significant ($p < 0.001$).

TABLE 1. Distribution of Examinees by Age in Both Study Groups. (Numbers in Parentheses are Percentages of the Same Study Group)

Age Groups (years)	Conventional n	AMHTS n
0–49	8 (10)	9 (12)
50–59	7 (9)	15 (20)
60–69	32 (41)	26 (34)
70 +	31 (40)	27 (34)
Total	78 (100)	77 (100)

Discussion

This paper addresses itself to a specific problem in delivery of health care in a public health organization: that of performing routine prehospitalization examinations for elective surgery in an efficient way. It has felt for a long time that the existing method for conducting these examinations, either during the hospital stay or by referring patients to existing (usually overloaded) laboratories and clinics, is an inefficient method that creates an extra burden on ambulatory services. Not only is it a time-consuming process for those who must be examined, but because of its fragmented nature, it also puts extra demands on physicians and medical staff in performing many routine tasks and much paper work. Moreover, it was found that in many instances the information gathered by these procedures was incomplete, not well organized and of variable quality. Thus,

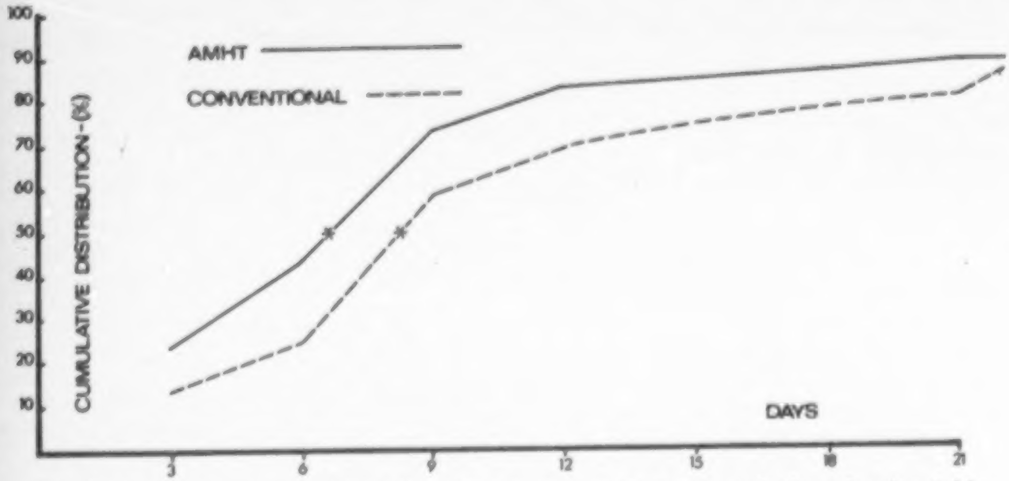


FIG. 2. Cumulative percentage distribution of the time duration from surgery to discharge from hospital for both groups.

new, more efficient (and tentatively more cost-effective) methods were explored. Utilization of multiphasic screening technology as a possible answer to this issue was investigated. A broad range of literature describing applications of AMHTS in compliance to various needs of ambulatory health care services[1-9] and, more specifically, as a part of hospital setups,[10,11] is available. AMHTS are particularly suitable for performing such uniform examinations, as is required for the concerned prehospitalization routine. Multiphasic screening units are dedicated systems geared for rapid and convenient processing of various populations. Most testing procedures are conducted by paramedical personnel and have built-in quality-control procedures. Standard concise reports are generated automatically and rapidly by the computer, and transferred to the appropriate ward. Addition-

TABLE 2. Distribution of Examinees by Diagnoses. (Numbers in Parentheses are Percentages of the Same Study Group)

Diagnoses	Conventional	AMHTS
Benign hypertrophy of prostate	36 (46)	28 (35)
Neoplasms of geniourinary organs	12 (15)	19 (24)
Malignant neoplasms of prostate	5 (6)	10 (13)
Urinary calculi	3 (4)	9 (11)
Other diseases of urinary system	9 (12)	2 (3)
Other diseases of male and female genital organs	6 (8)	5 (6)
Other ill-defined conditions	7 (9)	6 (8)
Total	78 (100)	78 (100)

ally, the same report also is sent to the personal physician (step, Fig. 1).

By referring patients for prehospitalization examinations to an AMHTS, several benefits could be demonstrated. The individuals who were examined saved a great deal of time (less than 4 hours vs. 2.1 days). Primarily, this is due to the fact that in the AMHTS track all tests were performed in one facility and controlled order (Fig. 1). Physicians were supplied with a printed, organized report, not fragmentary information that had been accumulated over a period of several weeks. Laboratories and staff of the health care organization were relieved of the time-consuming routine tasks involved in these examinations.

The finding that patients who were tested by the AMHTS had shorter postoperative stays in the hospital (by 1.8 days on the average) was not anticipated. Both study groups were quite similar in their age-sex-disease characteristics (Tables 1 and 2).

Similarly, no difference between the groups with respect to admission or operative diagnoses was found. Thus, none of these factors seemed to be responsible for the difference in the postoperative length of stay related to a possible bias affecting this outcome. Since there was no reason to assume that a postoperative stay would be affected by such basic and routine procedure, no specific questions in the evaluation questionnaire addressed themselves to this issue. This important finding (especially when substantial savings in hospitalizations costs are implied) requires further evaluation. By the same token, the feasibility of shortening preoperative hospital stay for AMHTS patients is now under investigation. This later issue seems to be primarily an administrative problem, requiring the introduction of some changes in scheduling admission and preoperative routines.

Results of this study support previous claims[1] that AMHT technology could be an efficient modality in solving particular needs in various segments of the health care system. Identifying such processes, where AMHT can be shown to provide more rapid, convenient or costeffective service (compared to the existing one) could lead to improvements in the delivery of care in many areas. Exploring new potential utilizations of AMHTS by different health care organizations also could assist in developing more versatile and extensive uses of this technology, therefore improving the overall efficiency and quality of the concerned health services.

References

1. Carel RS, Leshem G. Evaluation of the cost effectiveness of an automated multiphasic health testing system. Prev Med 1980;9:689.

2. Collen MF, ed. Multiphasic Health Testing Services. New York: John Wiley and Sons, 1978.

3. Dales LG, Friedman GD, Collen MF. Evaluating periodic multiphasic health checkups: a controlled trial. J Chronic Dis 1979;32:385.

4. Gilbert F, Nordyke R. Automated multiphasic health testing in multidisciplinary group practice. Prev Med 1973;1:261.

5. Fletcher SW, Sourkes M, Rabzel J, et al. Multiphasic screening, case finding tool in a teaching hospital medical clinic. JAMA 1977;237:887.

6. Holocomb FW. IBM's health screening program and medical data system. Journal of Occupational Medicine 1973;11:863.

7. Hsieh RKC. Multiphasic health testing in a public health service hospital. Med Instrum 1973;8:285.

8. Shapiro S, Frink R, Rosenberg C. A program to measure the impact of multiphasic health testing on health differentials between poverty and nonpoverty groups. Med Care 1972;10:207.

9. Yedidia A. California cannery workers program. Arch Environ Health 1973;27:259.

10. Davey PW, Letts HW, Anderson PA. Experience with a hospital laboratory admission screen. Can Med Assoc J 1970;103:140.

11. Durbridge TC, Edwards F, Edwards RB. An evaluation of multiphasic screening on admission to hospital. Med J Aust 1976;1:703.

BEST AVAILABLE COPY

Medical Care

September 1982
Vol. 20 No. 9
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Barriers to Medical Care of Mexican-Americans: The Role of Social Class, Acculturation, and Social Isolation

Racial/Ethnic Differences in Access to Health Care: Further Comments on the Use–Disability Ratio

Does Income Affect Mortality? An Analysis of the Effects of Different Types of Income on Age/Sex/Race-Specific Mortality Rates in the United States

Income and Race Differences in Children's Health in the Mid-1960s

Regional Differences in Hospital Use Standardized by Reported Morbidity

Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population?

Regression Analysis in Health Services Research: The Use of Dummy Variables

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Marion Athearn
Jennifer Gardon
Editorial Assistants

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year for individuals, $48.00, for institutions, $55.00; special price U.S. students, $35.00; all other countries except Japan, one year, for individuals, $59.00, for institutions, $68.00. Single copies $6.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, MD, and additional mailing offices.

September 1982
Volume 20, Number 9

CONTENTS

POLICY FORUM

Inequity in Access to Health Care: An Old and Persistent Worry. *Miriam Orleans* 881

ORIGINAL ARTICLES

Barriers to Medical Care of Mexican-Americans: The Role of Social Class, Acculturation, and Social Isolation. *Alan P. Chesney, Juan A. Chavira, Rogers P. Hall, and Howard E. Gary, Jr.* 883

Racial/Ethnic Differences in Access to Health Care: Further Comments on the Use–Disability Ratio. *Alfred C. Marcus and Jeffrey D. Stone* 892

Does Income Affect Mortality? An Analysis of the Effects of Different Types of Income on Age/Sex/Race-Specific Mortality Rates in the United States. *Jack Hadley and Anthony Osei* 901

Income and Race Differences in Children's Health in the Mid-1960s. *Linda N. Edwards and Michael Grossman* 915

Regional Differences in Hospital Use Standardized by Reported Morbidity. *Mark S. Blumberg* 931

Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population? *Noralou P. Roos and Leslie L. Roos, Jr.* 945

Regression Analysis in Health Services Research: The Use of Dummy Variables. *Lincoln Polissar and Paula Diehr* 959

COMMUNICATION

Assessing Needs for Community Hospital Continuing Medical Education. *Charles E. Osborne* 967

PUBLICATIONS RECEIVED 972

MEDICAL CARE
September 1982, Vol. XX, No. 9

Original Articles

Barriers to Medical Care of Mexican-Americans: The Role of Social Class, Acculturation, and Social Isolation

ALAN P. CHESNEY, PH.D.,* JUAN A. CHAVIRA, PH.D.,† ROGERS P. HALL, M.A.,‡
AND HOWARD E. GARY, JR., PH.D.§

Literature on medical care utilization by Mexican-Americans suggests that patterns of utilization can be explained, in part, by acculturation, social class and social isolation. The relative importance of these variables is explored in a survey of a rural Mexican-American community in Southern Texas. Data on 152 families are analyzed using a logit analysis, with utilization as the outcome variable. The results suggest that acculturation has a direct effect, while social class and social isolation interact to have a weak effect independent of acculturation. These results confirm the importance of understanding all three variables in planning and implementing medical care programs in Mexican-American communities.

RESEARCH on Mexican-American health care utilization may be categorized by the conceptual framework that guides the research. From an anthropologic framework, health behaviors are explained by referring to characteristics of the culture such as folk beliefs, acculturation, traditionalism and attitudes. Research conducted by Saunders,[1] Clark,[2] Rubel,[3] Welch,[4] and Farge[5] tend to support this explanation of health care utilization.

From a sociologic framework, health behaviors are explained in terms of the relationship of an individual or a family to the community. Such concepts as social isolation, social participation and social class have been related to utilization of health services. Research by Nall and Spielberg,[6] Bullough[7] and Hoppe and Heller[8] has supported the importance of sociologic factors in understanding health behaviors of Mexican Americans.

In some research, both conceptual frameworks have been used. Since traditionalism tends to be associated with isolation from the community, Reeder and Berkanovic[9] used these variables together to explain American health behavior. Social and cultural variables also have been combined by including both in the definition of a major study variable, such as social class.[10] None of these studies however, has analyzed the relative importance of cultural and social factors in studying health care utilization at a community level. Frequently, the cultural and social variables are linked so closely that it is difficult to separate them. Our study is directed to-

---

* Department of Psychiatry and Behavioral Sciences University of Texas Medical Branch, Galveston.

† School of Law, University of Texas, Austin.

‡ Department of Psychiatry and Behavioral Sciences University of Texas Medical Branch, Galveston.

§ University of Houston, Clear Lake City.

Data reported in this paper were collected pursuant to contract #232-78-0010 from the Bureau of Health Manpower, Department of Health, Education and Welfare.

Reprint requests: Alan P. Chesney, Department of Psychiatry & Behavioral Sciences, UTMB, Galveston, TX 77550.

0025-7079/82/0900/0883/$01.25 © J. B. Lippincott Co.

ward this problem. We attempt to identify the relative importance of social class, acculturation and social isolation in explaining health care utilization in a rural Mexican-American community.

Sociologists who have studied the relationship between minority groups and the dominant society stress the importance of the conceptual distinction between social class and culture.[11,12] Culture refers to norms and values which serve as guideposts or standards for behavior. Social structure, on the other hand, refers to the set of crystallized social relationships the members of society have with each other and with the society. This study preserves the conceptual distinction between cultural variables (cultural assimilation) and social structure variables (social class and social isolation). Such a distinction will enable us to explore the relative contributions of these variables to medical care utilization by Mexican-Americans.

Setting

The current study was conducted in Hidalgo, a small town in the Lower Rio Grande Valley of Texas. Hidalgo is located on the banks of the Rio Grande, 40 miles from the mouth of the river. The 2,500 residents of Hidalgo are predominantly Mexican-American (85 per cent) and most are poor. Approximately 60 per cent of the city's families live on less than $4,000 per year.

Hidalgo is connected to other towns in the Valley by Route 281, which ends at the International Bridge connecting Hidalgo with Reynosa, a large urban area with over 200,000 inhabitants. As a gateway to Mexico, Hidalgo serves as a point of entry into Mexico for nine million people per year. Since Hidalgo shares the revenue from the bridge, the town is in the unique position of public prosperity, despite the poverty of its residents. For example, the city has a modern police and fire station, it has a new high school building and all of its streets are paved.

Housing in Hidalgo is a study in contrasts: There are numerous houses built of plywood and cardboard that have dirt floors, there are mobile homes without wheels and there is an elegant mansion. An enumeration of dwelling units revealed that Hidalgo has 443 dwelling units, 53 per cent of which are in sound condition, 18 per cent are deteriorating and 28 per cent are dilapidated. The criteria for classifying the condition of a dwelling were developed by the United States Public Health Services in southern Texas.[13] The houses frequently are intermixed with sound houses. A few of the houses had unsanitary refuse storage (6 per cent), some had outhouses that could be observed from the road (13.5 per cent) and a small number of houses had livestock living in or near them.

At the time of this study, health services in Hidalgo were limited. There were no physicians or dentists practicing or living in Hidalgo. There were two lay midwives (Parteras) and a school nurse. Ambulance service to a Red Cross Clinic was available to Hidalgo residents. McAllen General Hospital, the nearest United States hospital, was 10 miles away. The Hidalgo County Health Department provided bimonthly clinics in Hidalgo that were mostly preventive. Consequently, many health problems remain undiagnosed and untreated.

Methods

Fifty per cent of the dwelling units were selected at random from the enumeration data. Of the 222 dwelling units included in the sample, 36 were vacant and 16 could not be located. The vacant dwelling units may have been migrant housing; the 16 units that could not be located may have been moved during the time between the enumeration and the survey. From the remaining 170 units, 152 complete interviews were obtained.

Data collection took place during the summer months of 1976. Each family was contacted by an interviewer who explained the purpose of the visit and requested cooperation. Interviews were conducted in English or Spanish, according to the respondent's preference. Only adult respondents were interviewed and, whenever possible, this person was the head of the household (either parent).

Social class of the family was operationally defined by family income. (In those cases where family income was unknown, social class was assigned on the basis of occupation.) If both income and occupation are unknown, then education was used to determine social class. These criteria allowed us to categorize each family as being in the upper or lower social class. The use of only two classes was considered appropriate since the community has a small middle class.

TABLE 1. Socio-demographic Characteristics of the 152 Mexican-American Families Interviewed

	N	%
Respondent characteristics		
Sex		
Male	15	9.8
Female	137	90.2
Age		
Under 25	18	11.8
25–34	41	27.0
35–44	23	15.1
45–54	39	25.7
55 and older	31	20.4
Years of formal education		
No formal education	17	11.2
Some, but less than 7 years	74	48.6
7–11 years	34	22.4
Completed high school	21	13.8
Education beyond high school	6	4.0
Family characteristics		
Number of persons in dwelling		
1–2	23	15.1
3–5	84	55.3
6–7	31	20.4
8 or more	14	9.2
Occupation of breadwinner		
Managerial–technical	11	7.2
Sales–clerical	11	7.2
Craftman–skilled labor	15	9.9
Semiskilled labor	22	14.5
Laborers and service workers	36	18.6
Unemployed, never worked, retired	57	37.5
Monthly family income		
Less than $86	21	13.8
$87–$250	29	19.1
$251–$499	38	25.0
$500–$999	31	20.4
$1,000 or more	15	9.9
Unknown	18	11.8

Social isolation was defined conceptually as the opportunity for social interaction, participation and support. This definition is consistent with definitions that have defined social isolation as one component of alienation.[7-14] In addition, this definition allowed us to examine social class and social isolation as separate variables. Previous measures of the social isolation component of alienation rely on the respondent's perception of his social network, e.g., "I can always count on my family for help when I need it."[13]

Meyers,[14] however, has suggested that the conceptual opposite of social isolation, social integration, may be measured more precisely by focusing on demographic and behavioral indicators. Therefore, the following demographic and behavioral data were combined into an index of social isolation: (1) marital status, (2) presence of a telephone in the dwelling unit, (3) amount of travel to Mexico, (4) amount of travel to other Valley towns and (5) length of time residing in Hidalgo.

Acculturation or cultural assimilation is conceptually defined in this research as "change of cultural patterns to those of the host society."[12] In this research, the term refers to acceptance of Anglo norms and values as the standard for judging one's behavior. Research on Mexican-Americans has documented the importance of language as an essential component of cultural assimilation. Olmedo[16] concluded that the major dimension of acculturation is language proficiency and preference. Olmedo's research also highlights place of birth as a major indicator of acculturation. It is used frequently as a measure of acculturation on the assumption that it reflects location of early socialization. However, since babies may be born in southern Texas and raised in Mexico, location of education has been included with location of birth and language proficiency to form an index of acculturation. The index indicates past value socialization and the present integration of the family into the norms and values of American (Anglo) society.

Data on utilization of health services were collected using two different methods. Respondents were asked the question, "If you or a member of your family is sick, what reasons would prevent you from seeing a doctor?" In addition, respondents were asked if a family member had experienced any of 20 specific symptoms in the past three months and if a physician was consulted if they had experienced them. The list of symptoms is an adaptation of Koos'[17] list. Two major problems emerged in using this list. One is that 27.6 per cent of the respondents reported no symptoms in their family and, therefore, would not be included in the analysis if symptom data were used to determine utilization behavior. The second major problem is that respondents showed a marked (75 per cent) tendency to report only their own symptoms rather than those of their family. An additional problem with the symptom data is that there is no measure of symptom severity. Therefore, the dependent variable is a predisposition to seek medical care given that a perceived need exists. The predisposition is measured by whether the respondent identifies barriers to medical care. This study does not attempt to define variables that are related to the definition of a problem as an illness.

Results

The mean size of the households in the survey was between four and five people. Small households of one or two persons generally were made up of older, retired persons. Single parent households, those in which the head of the household was separated, divorced or widowed, made up 19 per cent of the sample. The mean monthly income per household was approximately $420, 22 per cent of the households reporting either no income or refusing to provide the information. Employment information was collected on the head of the household. Thirty-nine per cent held skilled or semiskilled jobs, 24 per cent held unskilled jobs and 37 per cent were unemployed or retired. The average length of time at the present job for those currently working was 4.2 years. The cost of the housing computed from monthly rent or mortgage payments averaged $74 per month. Houses ranged in size from one to 12 rooms, with an average of approximately five rooms. Thirty-nine per cent of the housing would be classified as crowded by United States Census Bureau of Standards (1.1 or more persons per room). Over 90 per cent of the houses had stoves, refrigerators and radios, but only about half of the homes had running water or telephones.

The survey data on Hidalgo indicate that the sample selected is similar to the community as a whole. In addition, the sample appears representative of other communities along the Rio Grande. Comparison of our demographic descriptions with 1970 census demographic descriptions of the county provides striking similarities with regard to housing conditions, educational levels, income levels and length of residency at present location.

TABLE 2. Measures of the Level Acculturation for 152 Mexican-American Families

	N	%
Location of education		
Mexico	74	48.7
U.S.	78	51.3
Language spoken		
Spanish only	81	53.3
Spanish and English	71	46.7
Location of birth		
Mexico	70	46.1
U.S.	82	53.9
Summed acculturation score		
0,1 low	79	52.0
2,3 high	73	48.0

The first step in analyzing the data from this study is to develop indexes of social class, social isolation and acculturation. The second step is to identify the relationships among the independent variables. The third step is to identify the relative importance of the independent variables in explaining variations in the dependent

TABLE 3. Measures of Levels of Social Isolation for 152 Mexican-American Families

	N	%
Marital status		
Single, widowed, divorced	39	25.7
Married	113	74.3
Telephone in the dwelling unit		
No	76	50
Yes	76	50
Travel to Valley weekly		
No	67	44.1
Yes	85	55.9
Travel to Mexico weekly		
No	95	62.5
Yes	57	37.5
Length of time resided in Hidalgo		
Less than 5 years	67	44.1
More than 5 years	85	55.9
Summed social isolation		
0–3 high	63	41.4
3–5 low	89	58.6

TABLE 4. Distribution of Respondents ($N = 152$) Predisposition Toward Health Care Utilization by Social Class, Acculturation and Social Isolation

Social Class	Acculturation	Social Isolation	Utilization		Odds of Utilization
			No	Yes	
Low	Low	Low	15	14	0.93
Low	Low	High	11	6	0.55
Low	High	Low	7	6	0.86
Low	High	High	9	7	0.77
High	Low	Low	10	2	0.20
High	Low	High	13	8	0.62
High	High	Low	5	4	0.80
High	High	High	11	24	2.18

variable, regardless of the association among independent variables.

The measures of acculturation are reported in Table 2. The sample is evenly divided on each of the measures of acculturation and on the overall acculturation score. With regard to social isolation, it appears that the sample tends not to be isolated. Family members reported a tendency to travel to other towns in the valley and to have lived in Hidalgo for more than 5 years. These factors, in combination with a 75 per cent married rate, tend to push the social isolation scores toward low isolation.

TABLE 5. Chi-Square Summary Table for the Logits Analysis

Effects	df	χ^{2*}	p
Social class	1	0.01	0.94
Acculturation	1	4.11	0.04
Isolation	1	0.38	0.53
Social class × acculturation	1	2.34	0.29
Social class × isolation	1	3.41	0.06
Acculturation × isolation	1	0.07	0.78
Social class × acculturation × isolation	1	0.13	0.94

* χ^2 = likelihood ratio.

In response to the question, "If you or a member of your family were sick, what reasons would prevent you from seeing a doctor?", 53.9 per cent (81) stated that lack of money, transportation or both would stop them from seeing a doctor when they thought it necessary. In contrast, 26.1 per cent (71) reported that they would see a doctor when necessary. Those who identified barriers to physician utilization may be people who would delay in seeking medical care or people for whom medical care is not accessible.

To assess the effects of acculturation, social class and socialization on health service utilization, a logit analysis was performed using a log-linear model.[18–20]

This analysis permits the estimation of how each of the independent variables and their interactions affect the logit (i.e., the log of the odds) of utilization of health services. The contingency table for this analysis is presented in Table 4 and the chi-squared summary table is presented in Table 5. The results of this analysis indicate that only the main effect for acculturation is significant at 0.05 level.

From Table 4 we see that the odds of utilization for the low acculturation group are 0.61 while the odds for the high acculturation group are 1.28. Thus, increasing the level of acculturation from low to high approximately doubles the odds of health service utilization.

Since some researchers may be more comfortable with percentages than with odds, we have converted the estimated odds into percentages using the formula PERCENTAGE = 100 × (ODDS)/(ODDS + 1). Applying this formula to the estimated odds presented above, we find that 38 per cent of the low acculturation group would use health services, while 56 per cent of the high acculturation group would use these services.

A second interesting finding from these data is the suggestion that the effect of isolation on the odds of health service utilization depends on the person's social class. This is tested by the *social class × isolation* interaction in Table 5. This test fails to reach significance at the 0.05 level ($p = 0.07$), but we felt that the finding was sufficiently interesting to present here. Using the procedures outlined above, we estimated the odds of health service utilization separately for each combination of social class and social isolation. We found that for lower social class individuals, the estimated odds of utilization are 1.01 for the low isolation group (i.e., 50 per cent and 0.5 for the high isolation group (i.e., 33 per cent). For upper social class individuals, the odds are 0.788 for the low isolation group (44 per cent) and .854 for the high isolation group (46 per cent). Thus, for lower social class individuals, reducing isolation doubles the odds of health service utilization, while for upper class individuals, isolation essentially has no effect.

Discussion

The results of this community survey suggest two tentative conclusions in explaining medical care utilization by Mexican-Americans. One is that the acculturation is related directly to utilization. Those families that evidence more acculturation are more likely to utilize medical care. The second conclusion is that the impact of social isolation upon Mexican-American utilization is conditional upon social class and independent of acculturation.

This finding resolves the apparent conflict in the literature between a cultural and a social model explaining medical care utilization. In fact, while both cultural and social variables are necessary to explain community patterns of utilization, neither is sufficient to explain the variations in utilization that are described in Hidalgo. The results suggest a new direction for the health sciences research, which focuses on the relative contribution of social and cultural variables in explaining medical care utilization. In addition, this research suggests that analyzing variations in utilization patterns within one ethnic group will be a fruitful avenue of future research.

The conclusions suggested above are tentative because of the limitations of the research. It must be recognized that this survey was conducted in a Mexican-American community where medical care is relatively unavailable and inaccessible. In addition, the community offered what may be unique variations in acculturation and social isolation. Its proximity to Mexico and its population that has a high percentage of Mexican-Americans may have resulted in more families being identified as unacculturated and not socially isolated than would be true for Mexican-Americans nationally.

In addition to the question of generalizability, there is a problem in the precision of measurement. It is desirable to measure utilization of medical care with a prospective study of actual behavior. Use of hypothetical situations leaves a nagging question about its relevance to actual performance. In addition, more sophisticated measures of social class, social isolation and acculturation are in the process of being developed. The measures used in this research are consistent with other efforts to measure these sociologic concepts with demographic variables. More precise measurement will provide a more accurate test of the hypotheses proposed. In developing better instruments, it is important to remember that long structured interviews may not be the most useful method of gathering data about low income ethnic populations. Our experience suggests that a more conversational interview may be more effective.

Keefe and Casas[21] conclude that "Mexican-Americans who are Spanish-speaking, undereducated, and recent residents are for the most part unfamiliar with mental health services." The results of the study suggest that the statement also may be true for health services generally. However, our results suggest that within ethnic groups, variations regarding utilization are not simply a result of lack of knowledge about health care services. One explanation of this finding is that the combination of high income and a social network provide a family with alternatives to medical care. These alternatives may exist within the extended family. High income and high social isolation tend to increase the odds of utilization on the proportion of utilizers is not statistically significant, these trends suggest that higher class families who are socially isolated may lack alternative sources of care and, since they can afford medical care, they will seek it. Interventions which are aimed at increasing medical care utilization of low income, Mexican-Americans shoud focus on acculturation issues such as increasing their awareness of the rules and norms, which are an integral part of medical care. A strategy of reducing social isolation among low income Mexican-Americans as a means to increase medical care utilization is not aimed at the appropriate target group.

The relationship among social class, social isolation and medical care utilization is independent of acculturation. The conditional probability calculations on acculturation and utilization indicate that low acculturation is associated with low utilization. These results suggest that educational programs without regard to social class and social isolation may be beneficial. Since one aspect of acculturation is language, these results suggest that increasing the number of Spanish-speaking medical personnel could result in increased medical care utilization by Mexican-Americans.

These results suggest that the tendency to utilize or not utilize medical care involves a complex social and cultural process. Intervention strategies which oversimplify the process of seeking health care by focusing on either social class, social isolation or acculturation will not produce dramatic changes in utilization rates. We have explored the interaction between these variables in order to document the importance of various population subgroups within the Mexican-American population. Two subgroups tend toward more utilization of medical services. Each group will require unique intervention strategies aimed at increasing medical care utilization.

Future research in this area should take advantage of recent conceptual and methodologic developments in utilization behavior and acculturation. The symptoms–response ratio developed by Anderson and Smedbey[22] offer a more sophisticated measure of underutilization of medical care services. That technique compares a ratio of visits per symptom to the expected ratio of visits per symptom as defined by a panel of physicians. This measure has at least two advantages. First, it eliminates the problem of symptom severity from the equation. Second, it focuses the research on group rates rather than individual behaviors.

Subsequent to our field work, an acculturation scale for Mexican-Americans has been developed by Cueller.[23] The use of an acculturation scale coupled with demographic characteristics (place of birth, place of education, etc.) should provide a powerful measure of acculturation.

The relationship among acculturation, social class, social isolation and medical care utilization proposed here have implications for delivery of health services to rural Mexican-American communities. First, health planners must recognize the diversity that exists among Mexican-Americans. It is not sufficient to develop programs that are based on simplistic notions about "the Mexican-American" population. Some portions of a Mexican-American community find medical care readily accessible. Other segments find the health care system to be unaccessible for very different reasons. Programs geared to increase accessibility of services through education interventions will probably be effective among the segment of the community that is relatively poorly acculturated. However, educational programs are not sufficient; programs which attempt to provide social support mechanisms to increase accessibility will need to be developed carefully. In some cases, the interventionist should link into an existing support system; in other groups a social support system needs to be created. Determining what type of intervention should be implemented to increase health care accessibility in a community requires knowledge of the levels of acculturation, social class and social isolation in the community. Future needs assessments of health care in Mexican-American communities should consider these variables.

Acknowledgment

The authors extend their appreciation to Dr. W. W. Schlottstaedt and the anonymous reviewers for their contributions to this manuscript.

References

1. Saunders L. Cultural differences and medical care. New York: The Russel Sage Foundation, 1954.

2. Clark M. Health in the Mexican-American culture. Los Angeles: University of California Press, 1959.

3. Rubel A. Across the tracks: Mexican-Americans in Texas City. Austin: University of Texas Press, 1966.

4. Welch S, Comer T, Steinman M. Social and attitudinal correlates of health care among Mexican-Americans. J Health Soc Behav 1973;14:212.

5. Farge EJ. La Vida Chicana: Health care attitudes and behaviors of Houston Chicanos. San Francisco: Rand E. Research Associates, 1975.

6. Nall FC, Spielberg T. Social and cultural factors in the responses of Mexican-Americans to medical treatment. J Health Soc Behav 1967;9:299.

7. Bullough B. Poverty, ethnic identity, and preventive health care. J Health Soc Behav 1972;13:347.

8. Hoppe SK, Heller PL. Alienation, familism, and the utilization of health services by Mexican-Americans. J Health Soc Behav 1975;16:304.

9. Reeder L, Berkanovic E. Sociological concommitants of health orientations: A partial replication of Suchman. J Health Soc Behav 1973;14:134.

10. Madsen W. The Mexican-Americans of South Texas. New York: Rinehart & Winston, 1964.

11. Schermerhorn RA. Comparative ethnic relations: a framework for theory and research. New York: Random House, 1970.

12. Gordon M. Human nature, class, and ethnicity. New York: Oxford University Press, 1978.

13. Waldrop RH. Socioeconomic stratification in community health. Washington, DC: U.S. Department of Health, Education, and Welfare 1971:1.

14. Dean D. Alienation: Its meaning and measurement. Am Sociol Rev 1961;24:753.

15. Meyers J, Lindenthal J, Pepper M. Life events, social integration and psychiatric symptomatology. J Health Soc Behav 1975;16:421.

16. Olmedo EL, Padilla AM. Empirical and construct validation of a measure of acculturation for Mexican-Americans. Journal of Soc Psycho 1978;105:179.

17. Koos EL. The health of Regionville. New York: Columbia University Press, 1954.

18. Bishop YMM. Full contingency tables, logits, and split contingency tables. Biometrics 1969;25:383.

19. Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis. Cambridge, Mass.: The MIT Press, 1975.

20. Swafford P. Three parametric techniques for contingency table analysis: a non-technical commentary. Am Sociol Rev 1980;45:664.

21. Keefe S, Casas J. Mexican-Americans and mental health: a selected review and recommendations for mental health delivery. Am J Community Psychol 1980;8:303.

22. Anderson R, Smedbey B. Changes in response to symptoms of illness in the United States and Sweden. Inquiry 1975;12:116.

23. Cueller I. The underutilization of mental health facilities and services by Mexican-Americans: who, where, how much, and why? Proceedings of the Texas-New Mexico Symposium on the Delivery of Mental Health Services to Mexican-Americans, 1979.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Racial/Ethnic Differences in Access to Health Care: Further Comments on the Use–Disability Ratio

ALFRED C. MARCUS, PH.D.,[*] AND JEFFREY D. STONE, PH.D.[†]

Indicators of access to health care generally fall into one of two broad groups. The first group of indicators focuses on factors responsible for differential rates of use and often are assessed through reports of patient satisfaction. The second group of indicators focuses more specifically on actual rates of use, and therefore represents a more objective measure of access. One of the most widely used objective indicators of access is the use–disability ratio. This article illustrates some problems in using this ratio in health services research and some strategies one might adopt to minimize these problems. The data reported below were obtained from an analysis of racial/ethnic differences in use of services that was conducted as part of the 1979 Los Angeles Health Survey (N = 1003). Our findings suggest that 1) blacks had greater access to health care than either white/Anglos or Hispanics, particularly for respiratory and musculoskeletal problems, and 2) the use–disability ratio should be used cautiously, since it is sensitive to subgroup differences in the chronicity and cause of the disability.

PERHAPS no concept is as central to health services research as patient access to health care. Ironically, few concepts also are as ill-defined.[1] In an attempt to rectify this problem, Penchansky and Thomas[2] recently introduced a taxonomy of access that focuses on factors hypothesized to influence use of services. According to their model, access is "a general concept that summarizes a set of more specific dimensions describing the fit between the patient and the health care system." The specific dimensions of access identified by Penchansky and Thomas included availability, accessibility, accommodation, affordability and acceptability.

Although the degree of fit between the individual and the health care system, as determined by reports of patient satisfaction, is important to understanding health services utilization, there is also a need to develop objective measures of access that focus on the frequency of using health services.[3] For example, recent data from the National Center for Health Statistics show that people with lower incomes have more contact with doctors per year than people with higher incomes.[4] However, this difference does not mean that low income groups have greater access to health care than middle and upper income groups, since it might also reflect variations in the *need* for health care. As both Aday[4,5] and

---

[*] Director of Program Evaluation, UCLA Jonsson Comprehensive Cancer Center, Division of Cancer Control, University of California, Los Angeles.

[†] Research Associate, School of Public Health, University of California, Los Angeles.

Data reported in this paper were collected pursuant to grant number 5-R18-CA-18451 "Processes in Health Behavior and Cancer Control," awarded by the National Cancer Institute to the late Leo G. Reeder, Principal Investigator.

Requests for reprints: Alfred C. Marcus, Ph.D., Director of Program Evaluation, UCLA Jonsson Comprehensive Cancer Center, Division of Cancer Control, 10920 Wilshire Boulevard, Suite 1106, Los Angeles, California 90024.

0025-7079/82/0900/0892/$01.25 © J. B. Lippincott Co.

Andersen[5] observed, access to health care is best considered within the context of whether people who actually need medical care receive it. It is in defining and assessing this need that health status indicators become essential to the measurement of access. .

One of the most widely used objective indicators of access is the use–disability ratio, which is defined as the number of doctor visits per 100 disability days. Thus, this ratio combines the features of an epidemiologic rate with a measure of use that has been standardized on the basis of need (i.e., disability days). This ratio made its first appearance in an article by Aday,[4] and since then has been reported in several other national, state and community surveys.[6–10]

Although the use–disability ratio has been used extensively in several prominent studies, it has also had its share of critics. For example, Fox and Bice[11] criticized the index because it assumes a linear relationship between need for care and disability days. Chen[3] has raised two other objections. First, the index has no objective norm or standard, making it difficult to interpret from a policy perspective (e.g., does a higher ratio indicate more appropriate use of health services or overutilization of health services?). Second, Chen notes that the cause of the disability is not taken into account in this index. Consequently, in comparing rates among population subgroups, one must assume that a comparable number of disability days represents a comparable need for health care—regardless of the cause of the disability. Montiero[8] made a similar argumeant in her report on the Rhode Island Health Survey. She criticized the ratio because it failed to distinguish between acute and chronic illness. That is, if the source of the disability is an acute illness, then a visit to the physician may be more likely than if the illness was long-standing or chronic. Accordingly, "since economic levels and age groups are likely to differ in the underlying source of restricted activity days, then differences in the use/need ratios for these groups may to some extent reflect the type of illness experienced rather than any real difference in access to medical care."

In this article, we report findings from a comparison made between the use–disability ratio and another indicator of access. Our purpose is to illustrate some of the problems that may be encountered in using this ratio in health services research and some of the strategies one might adopt to minimize these problems.

Methods and Procedures

The data reported were collected during the summer of 1979 as part of the Los Angeles Health Survey (LAHS). Fieldwork was conducted by the UCLA Survey Research Center of the Institute for Social Science Research (ISSR). ISSR maintains a sampling frame containing approximately 20,000 computer-readable addresses sampled from Los Angeles County on an area probability basis. Samples drawn from the frame may be characterized as three stage samples with probabilities proportional to size.[12] One adult per household was selected to be interviewed using the Kish sampling table. From the 1,531 eligible respondents, 1,003 completed interviews were obtained (65.5 per cent) consisting of 63.4 per cent white/Anglos, 12.3 per cent blacks and 16.7 per cent Hispanics. Further details of the sample demographic characteristics are reported by Frerichs, et al.[13] ‡

The measures of restricted activity and utilization examined are similar to those used in the Health Interview Survey. The number of restricted activity days represents the total number of bed days and reduced activity days reported during the past two months. Thus, the number of restricted activity days ranged from 0 to 60. Information also was obtained about acute and chronic illness. For acute illness, the questionnaire item asked: "Thinking back to the past two months, were you sick or did you have any accidents or injuries at any time during the past two months? Please do not include any recurrent or continuing health problems that you might have all of the time." For chronic illness, the questionnaire asked: "Do you have a chronic, that is, a recurring or continuing health problem or condition; a health problem that you have all of the time or that makes you ill from time to time?" Up to three problems were recorded for both acute and chronic conditions. For each problem reported, respondents were asked to describe the medical problem and the number of times they contacted a doctor or other health professional about the problem during the past two months. Each reported symptom or problem was coded using the National Ambulatory Medical Care Survey Symptom Classification.[14]

---

‡ In this analysis, respondents were classified as Hispanic if they identified themselves a being Chicano or Mexican descent. Frerichs, et al.[13] also included in this category those who did not classify themselves as being of Chicano or Mexican descent, but who did classify themselves as being of "Other Spanish descent" (e.g., Cuban, Puerto Rican, etc.).

Measures of Access

In this analysis we have compared two of the leading "objective" indicators of access: the use–disability ratio as originally introduced by Aday,[4] and a measure of access similar to the one proposed by Chen.[3] The use–disability ratio is an aggregate measure calculated by dividing the number of physician contacts during a specified time period by the total number of disability days reported during the same period:

$$\text{Use–disability ratio} = \frac{\sum_{i=1}^{n} \text{MD}}{\sum_{i=1}^{n} \text{DD}}$$

where: n = number of persons in group reporting one or more disability days;
MD = number of doctor contacts;
DD = number of disability days (reduced activity days or bed days).

Since use–disability ratios are calculated for groups rather than individuals, it is difficult to test them statistically.§ Thus, inferences about "significant" differences on the use–disability ratio are usually based on a subjective evaluation as to what constitutes a meaningful difference (e.g., a difference of 5+ doctor contacts per 100 disability days).

The measure of access proposed by Chen is defined as the "proportion of people reporting a sign or symptom who acutally visited a physician":

$$p = \frac{n}{N}$$

where: n = number of persons reporting a sign or symptom who visited a physician;
N = number of persons reporting a sign or symptom.

At the individual level, scores on this measure can either be 0 or 1, depending on whether the person had symptoms and visited a physician (score = 1), or had symptoms and did not visit a physician (score = 0). Since this index uses the individual rather than the group as the statistical unit of analysis, significance testing is easily accomplished (e.g., by using χ^2 or F-statistics).

As originally proposed by Chen, the denominator of the index (i.e., the number of people reporting a serious sign or symptom) would be calculated from a predetermined list of health problems requiring medical attention. These problems would be identified by a panel of physician judges and included in the questionnaire in the form of a checklist. All respondents then would be asked whether they experienced a sign or symptom on the list and whether they saw a doctor about that problem.

---

§ The closest one can come to a direct statistical test of group level ratios involves using approximation methods like those discussed by Cochran.[15]

Since the LAHS was not specifically designed to examine Chen's index, we are not able to replicate it exactly as prescribed. In particular, our symptom data were not coded in sufficient detail to permit medical evaluations of their severity or appropriateness for medical care. Nonetheless, we have examined a variant of this index, which is defined as the proportion of people with a particular health problem that contacted a doctor.[‖]

Results

Racial/Ethnic Differences in Access

Criticisms of the use–disability ratio have focused on the use of disability days as a measure of need and the fact that the ratio does not take the cause of the disability into account. Both of these problems became evident in our analysis of racial/ethnic differences in access to heal care. For example, Table 1 reports racial/ethnic differences on our two indicators of access: the use–disability ratio and the proportion of people with health problems who saw a doctor. As shown, blacks scored higher than white/Anglo and Hispanic respondents on both measures, suggesting that blacks had relatively greater access to health care. Notice, however, that while Hispanics scored lower than white/Anglos on the use–disability ratio by a factor of 2.5 doctor visits, their scores were essentially identical on the other measure of access.

This difference for the Hispanic subgroup illustrates an often-discussed problem with the use–disability ratio, namely, its sensitivity to subgroup differences in long-term disability. For example, 12.8 per cent of the Hispanics reported 60 days of restricted activity during the past two months, compared with only 8.3 per cent for white/Anglos and 9.8 per cent for blacks. Similarly, 14.5 per cent of the white/Anglos and 12.2 per cent of the blacks reported 30 or more restricted activity days, compared with 23.1 per cent of the Hispanics. This difference in long-term disability, of course, can disproportionately increase the denominator of the ratio for Hispanics—thereby reducing their overall use–disability ratio.

One strategy to minimize the effects of statistical outliers involves transforming the variable disability days (e.g., log transformation) before calculating the use–disability ratio. In so doing, we found that blacks were ranked highest (ratio using log disability days = 1.47), with white/Anglos (ratio = 1.31) and Hispanics (ratio = 1.34) recording somewhat lower scores. It is of interest to note that this ranking corresponds perfectly to that obtained from the other measure of access.

Another strategy to minimize the effects of outliers involves trimming or truncating the number of disability days at various cut-points. For example, one could truncate the number of disability days reported at 30+ so that the maximum number any one person can report would be 30. As shown in Table 2, we found very little difference among racial/ethnic subgroups when the scores were trimmed at 22+ days. However, as the number of disability days is allowed to increase (i.e., as we move from left to right in Table 2), racial/ethnic differences increase, with Hispanics showing progressively less access relative to their disability than either white/Anglo or black respondents. This pattern in the ratios indicates that the magnitude of the racial/ethnic differences are

---

[‖] We also examined the proportion of health problems reported by an individual that resulted in a doctor contact. However, since most people in this survey reported a maximum of one health problem during the two month recall period, this variable was virtually identical to the one reported above (i.e., the proportion of people with health problems who contacted a doctor).

TABLE 1. Comparison of Two Measures of Access by Race/Ethnicity

	Use–Disability Ratio (past two months)*		Proportion of People With Health Problems That Reported a Doctor Contact (past two months)†	
Sample	20.3	(336)	0.52	(573)
Race/Ethnicity				
White/Anglo	20.7	(228)	0.49	(395)
Black	22.9	(41)	0.69	(64)
Hispanic	18.3	(39)	0.47	(73)

* Use/disability ratio × 100 equals the number of doctor contacts per 100 restricted activity days. This ratio is calculated only for those respondents who had at least one restricted activity day (N = 336).

† Calculated only for those people who reported at least one chronic or acute health problem during the past two months (N = 573). The denominator of the proportion equals the number of people with one or more health problems.

not uniform—partly because the subgroups differed in their rates of long-term disability.

The source of this problem became even more apparent upon closer examination of the data. That is, there were three Hispanics in this study reporting a total of 156 restricted activity days (60 + 60 + 36) and no doctor contacts. Neither the white/Anglo nor black subgroups had a similar subset of outliers. When we deleted these three people from the analysis, the use–disability ratio for Hispanics increased to 24.8, ranking them first among the three racial/ethnic groups. This particular ranking, however, illustrates a second problem with the use–disability ratio; namely, the failure of this ratio to consider differences in the etiology of the disability.

As column A in Table 3 shows, three major symptom categories were most likely to trigger a multiple doctor visit during the two-month recall period: acute and chronic musculoskeletal problems and cardiovascular problems. Moreover, as shown in column B, although proportionately fewer Hispanics with reduced activity reported cardiovascular problems, they did report a relatively high rate of acute musculoskeletal problems, and, by far, the highest rate of chronic musculoskeletal problems. Interestingly, when multiple doctor contacts were deleted from the numerator of the ratio, the obtained racial/ethnic differences were consistent with those found on the other indicator of access. That is, blacks were ranked first (ratio = 14.1), followed by Hispanics (11.1)

TABLE 2. Comparison of Use/Disability Ratios After Truncating Individual Reports of Restricted Activity Days by Race/Ethnicity*

	Trimmed 22+ Days	Trimmed 30+ Days	Raw Score
Race/Ethnicity			
White/Anglo (N = 228)	32.6	27.7	20.7
Black (N = 41)	33.7	30.5	22.9
Hispanic (N = 39)	31.8	26.2	18.3

* Use/disability ratio × 100 equals the number of doctor contacts per 100 restricted activity days. This ratio is calculated only for those respondents who had at least one restricted activity day during the past two months (N = 336).

and white/Anglos (10.9) with virtually identical scores. Thus, it would seem that Hispanics scored higher on the use/disability ratio (after deleting the three outliers), in part, because they reported more multiple doctor contacts than either black or white/Anglo respondents.

In the above analyses, the use–disability ratio was found to be sensitive to subgroup differences in long-term disability that could inflate the denominator of the ratio, and to differences in the etiology of the disability that could produce differential reports of multiple doctor contacts. The other indicator of access also has limitations, mainly because it uses health problems as the measure of need without taking into account variations in the severity of the problem. Thus, to examine the robustness of the racial/ethnic differences reported in Table 1, we compared these subgroups within categories of disability. As shown in Table 4, all of the comparisons continued to show greater use of health services by blacks than either white/Anglo or Hispanic respondents. Indeed, among the subgroups reporting 7+ restricted days, all 21 blacks made contact with a doctor, compared to 67 per cent and 79 per cent for white/Anglos (N = 81) and Hispanics (N = 14).

Up to this point, two well-known objective indicators of access have been compared: the use–disability ratio and the proportion of people with symptoms who contacted a doctor. A third index, defined as the proportion of people reporting disability days and had contacted a doctor, was also examined in this analysis. Using this index as a guide, blacks are ranked again as having the highest access, with 75.6 per cent reporting a doctor contact, compared with only 56.4 per cent for Hispanics and 52.2 per cent for white/Anglos ($p < 0.05$).

In yet another analysis, multiple regression techniques were used to estimate the number of doctor contacts reported by each ethnic group after statistically control-

TABLE 3. Proportion of People Reporting Particular Symptom Codes That Resulted in a Multiple Doctor Contact, and the Proportion Reporting These Symptoms by Race/Ethnicity

Symptom Codes†	A Proportion of Sample That Reported a Multiple Doctor Contact for Symptom		B Percentage of Racial/Ethnic Group With Reduced Activity That Reported Symptom*		
	%	N	White/Anglo	Black	Hispanic
Flu (313)	6.9	72	20.2	14.6	15.4
Acute respiratory (300–399)	9.3	118	29.8	24.4	17.9
Acute musculoskeletal (400–499)	35.4	65	9.2	22.0	20.5
Chronic musculoskeletal (400–499)	25.8	132	21.9	19.5	30.8
Cardiovascular (200–299)	34.3	102	13.6	12.2	7.7

---

* Percentages are based on number of people in group that reported at least one restricted activity day, and thus were included in the use/disability ratio. The number of white/Anglos was 228; blacks, 41; and Hispanics, 39.

† Symptoms were coded using the National Ambulatory Medical Care Survey Symptom Classification. Symptom codes are shown in parentheses.

TABLE 4. Proportion of People With Health Problems Who Made Contact With a Doctor by Race/Ethnicity and Level of Restricted Activity

	Proportion of People With Health Problems Reporting a Doctor Contact*					
	1+		3+		7+	
Race/Ethnicity	Restricted Days		Restricted Days		Restricted Days	
White/Anglo	0.52	(228)	0.61	(137)	0.67	(81)
Black	0.76	(41)	0.93	(29)	1.00	(21)
Hispanic	0.56	(39)	0.66	(29)	0.79	(14)
Significance level	$p < 0.05$		$p < 0.01$		$p < 0.01$	

* Number in parentheses represents the denominator of the percentage.

ling for differences in reports of illness and disability. As shown in Table 5, when no adjustments are made for racial/ethnic differences in reports of symptoms and disability, blacks averaged 1.47 doctor contacts, followed by white/Anglos (1.13) and Hispanics (0.76). When racial/ethnic differences in reports of symptoms and disability were controlled statistically, blacks were ranked first again—but by a somewhat larger margin ($p < 0.10$). As an aside, we might note that several cross-product interaction terms were also examined in this analysis, including race/ethnicity (dummy coded) by disability days. However, none of these interaction terms was statistically significant.

Although the number of cases was small, an examination of individual symptom codes showed that blacks were more likely to visit a doctor for respiratory and musculoskeletal symptoms (Table 6). A closer examination of these data, however, failed to uncover a common denominator that might explain this difference. For example, in one analysis we grouped all respondents reporting either respiratory or musculoskeletal problems (N = 278). Dummy variables representing blacks and Hispanics then were included in a multiple regression containing several other predictors, including respondent age, sex, total family income, educational and occupational ranking, number of restricted activity days in the past two months, regular source of health care (1 = no regular person or place, 2 = regular place but no regular person, 3 = regular person), and health in-

TABLE 5. Estimated Number of Doctor Contacts During the Two-Month Recall Period by Race/Ethnicity (N = 927)*

	White/Anglo	Black	Hispanic
Unadjusted scores	1.13	1.47	0.76
Adjusted scores	1.06	1.68†	1.36

* Number of doctor contacts was estimated using multiple regression. Race/ethnicity was dummy coded, with white/Anglo serving as the comparison group. The unadjusted scores represent the predicted scores for the three racial/ethnic groups from a regression containing only the two racial/ethnic variables (black vs. white/Anglo and Hispanic vs. white/Anglo) and the constant (white/Anglos). The adjusted scores represent the predicted scores from a regression containing the two racial/ethnic variables plus the number of restricted activity days (log of raw score + 1) and illnesses reported during the same recall period (i.e., the past two months). To estimate values for these latter two terms in the equation, their respective mean scores were used.

† $p < .10$.

TABLE 6. Proportion of People Who Contacted a Doctor About a Particular Symptom by Race/Ethnicity

	Proportion of People With Problem Who Made a Doctor Contact (past two months)		
System Category*	White/Anglo	Black	Hispanic
Respiratory (300–399)	33.7 (89)	54.5 (11)	30.0 (10)
Acute musculoskeletal (400–499)	52.4 (42)	83.3 (12)	70.0 (10)
Chronic musculoskeletal (400–499)	35.6 (87)	50.0 (14)	25.0 (20)
Cardiovascular (200–299)	56.0 (75)	71.4 (14)	75.0 (8)

Number in parentheses represents the denominator of the percentage.

* Symptoms were coded using the National Ambulatory Medical Care Survey Symptom Classification. Symptom codes are shown in parentheses.

surance (0 = does not have insurance 1 = has insurance). Results from this analysis showed that, controlling for all other variables in the equation, blacks were still significantly more likely to report a doctor visit than white/Anglo respondents ($\beta = 0.14$, $p < 0.05$). There were no significant differences between Hispanic and white/Anglo respondents ($\beta = 0.05$, NS).

Summary and Discussion

Two potential problems with the use–disability ratio were illustrated in these analyses. The first problem is that the ratio is sensitive to subgroup differences in chronic disability. Thus, when using this ratio for comparative purposes, it will be important to determine whether the subgroups differ in their rates of long-term disability. If the subgroups differ in this regard, then some type of adjustment should be made (e.g., "trimming" the distribution of disability days, deleting outliers, using a log transformation, etc.).

The second problem is that the ratio does not consider subgroup differences in the cause of the disability. Certain sources of disability (e.g., acute vs. chronic illness) may be more or less likely to result in a doctor contact. Moreover, once contact with a doctor has been made, certain sources of disability may be more or less likely to result in multiple doctor contacts (e.g., cardiovascular vs. respiratory problems). If the sample size is large enough, it therefore may be advisable to examine use–disability ratios separately for acute vs. chronic illness, or perhaps within categories of symptoms. Of course, making such comparisons presumes that one can link reports of restricted activity days with reports of doctor visits for particular symptoms.

Given these problems with the use–disability ratio, one might question the value of using this index at all. We found the use–disability ratio to be somewhat awkward to use, since a considerable amount of preliminary work had to be done to determine whether nonaccess factors were influencing the results. However, if subgroups are comparable in their illness profiles (e.g., acute vs. chronic illness, reports of particular symptom codes, etc.) and rates of long-term vs. short-term disability, or are made comparable by "trimming," deleting outliers, etc., then the use–disability ratio would seem to represent a reasonable measure of access. That is, all things being equal, subgroups reporting higher rates of use relative to their disability may indeed have greater access than subgroups scoring lower on the use–disability ratio. Yet, perhaps the most serious limitation of the use–disability ratio concerns the unit of analysis. Because the use–disability ratio is a group level variable, subgroup differences are not easy to test. Indeed, of all the analyses we conducted, the results obtained from the multiple regression were the most informative—offering significance testing with multivariate controls, and estimates of change in the dependent variable (i.e., number of doctor visits) attributable to other measured variables of interest (e.g., disability days, income, race/ethnicity, etc.).

Finally, the inability to explain why blacks were more likely to report a doctor contact for respiratory and musculoskeletal problems should be mentioned. This difference occurred independently of standard sociodemographic variables (e.g., age, sex, income, etc.), as well as the more policy-relevant variables pertaining to regular source of care and health insurance. Perhaps the answer lies in racial/ethnic differences in the dimensions of access identified by Penchansky and Thomas[2] (i.e., availability, accessibility, accommodation, affordability and acceptability), or in the type of health-related advice offered by friends, family, spouse, etc.[16–19] This ambiguity illustrates the need to link objective indicators of access with factors hypothesized to influence use of services. Thus, in the final analysis, we again are reminded that informed policy decisions require an assessment of use relative to need, as well as an understanding of the variables responsible for the differential rates of use. Given these two requirements, it would seem that the most useful indicators of access will be those that make the individual, rather than the group, the main focus of the analysis.

References

1. Aday LA, Andersen R. Access to medical care. Ann Arbor, MI: Health Administration Press, 1975.

2. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. Med Care 1981;19:127.

3. Chen MK. Measuring need for health services: a proposed model. Med Care 1979;17:210.

4. Aday LA. Economic and noneconomic barriers to the use of needed medical services. Med Care 1975;13:447.

5. Aday LA. Letter to the editor. Med Care 1976;14:717.

6. Andersen R. Health status indices and access to medical care. Am J Public Health 1978;68:458.

7. Aday LA, Andersen R, Fleming GV. Health care in the United States: equitable for whom? Beverly Hills: Sage Publications, 1980.

8. Montiero LA. Monitoring health status and medical care. Cambridge: Ballinger Publishing Company, 1976.

9. Marcus AC, Reeder LG, Jordan LA, Seeman TE. Monitoring health status, access to health care, and compliance behavior in a large urban community: a report from the Los Angeles Health Survey. Med Care 1980;18:253.

10. Nathanson C. Social roles and health status among women: the significance of employment. Soc Sci Med 1980;14A:463.

11. Fox PD, Bice TW. Socioeconomic status and use of physician services revisited. Med Care 1976;14:714.

12. Sumner J. The 1976 LAMAS frame and master sample: technical description. Los Angeles: Institute for Social Science Research, 1976.

13. Frerichs RR, Aneshensel CS, Clark VA. The prevalence of depression in Los Angeles County. Am J Epidemiol 1981;113:691.

14. National Center for Health Statistics. The national ambulatory medical care survey: symptom classification. USDHEW, Public Health Service, Vital and Health Statistics, U.S. Department of Health, Education and Welfare, Series 2:63, 1974. (DHEW Publication No. 74-1337.)

15. Cochran WG. Sampling techniques. New York: John Wiley and Sons, 1977.

16. McKinlay JB. Some approaches and problems in the study of the use of services: an overview. J Health and Soc Behav 1972;13:115.

17. Langlie JK. Social networks, health beliefs and preventive health behavior. J Health Soc Behav 1977;18:244.

18. Safer MA, Tharps, QT, Jackson TC, Leventhal H. Determinants of three stages of delay in seeking care at a medical clinic. Med Care 1979;17:11.

19. Berkanovic E, Telesky C, Reeder S. Structural and social psychological factors in the decision to seek medical care for symptoms. Med Care 1981;19:7.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Does Income Affect Mortality? An Analysis of the Effects of Different Types of Income on Age/Sex/Race-Specific Mortality Rates in the United States

JACK HADLEY, PH.D.,* AND ANTHONY OSEI, M.A.†

This article explores the question of whether or not higher incomes are associated with lower mortality rates. Some recent research on this issue has suggested that income either has no effect on or may even be positively correlated with mortality rates. By contrast, earlier studies consistently found a negative relationship—higher income (or economic status) was generally associated with lower mortality rates. This paper extends the prior research in two significant ways. First, the issue is analyzed separately for eight adult and four infant age/sex/race-specific population cohorts. Second, total family income is broken down into several components to investigate whether different types of income have differential effects on mortality rates. In addition, the problem of untangling the joint effects of education and income on mortality also is explored. The results tend to support the hypothesis that higher income is associated with lower mortality rates. However, the magnitude of the impact of income is small, although it is consistently larger for infants than for adults.

ALL COUNTRIES VALUE good health as an important part of individual welfare. It is also a significant factor in the level of economic activity, influencing both labor force participation and worker productivity. One variable that is used frequently to measure health is the mortality rate. It generally is agreed that the recent decreases in mortality rates in the United States have been due to several factors: scientific and technologic advances that increased the supplies of commodities contributing to good health; greater educational attainment, making people better health "producers"; increased use of improved quality medical care; and higher real incomes. For example, more than 15 years ago the American Medical Association stated, "Medical science does not seek major credit for the improvements in the health level during the past 25 years. Certainly, our standards of living and higher education levels have contributed substantially to the betterment of the health level in the United States."[1]

Such statements partially explain why health services researchers recently have tended to focus on socioeconomic conditions to explain mortality differentials. This article continues to work in this vein

---

* Senior Research Associate, The Urban Institute, Washington, D.C.

† Research Assistant, The Urban Institute, Washington, D.C.

Supported by Grant No. 5-R01-HS02790 from the National Center for Health Services Research, OASH, USDHHS.

The findings and opinions expressed in this paper are solely those of the authors and do not necessarily reflect the positions of The Urban Institute or the Department of Health and Human Services.

0025-7079/82/0900/0901/$01.50 © J. B. Lippincott Co.

by focusing on the relationship between income and mortality rates. Specifically, is income a significant factor in explaining geographic variations in age- sex- race-specific mortality rates? This question is particularly interesting in this time of "budget crisis." With all levels of government concerned about the size and growth of transfer payments, which constitute a substantial portion of the incomes of certain population subgroups, it is important to ask what consequences cuts in transfer payments might have on the health of the population. Will mortality rates increase, decrease or level off with changes in income?

The existing evidence on this issue is conflicting. Some studies imply that death rates tend to be correlated negatively with income, while others postulate an insignificant or even positive correlation. Although the factors underlying the latter relationship are not well understood, several hypotheses have been offered to explain this apparent paradox. They include the existence of a ceiling on the ability of standards of living to improve health which already has been reached by certain population groups; positive income-elasticities of demand for certain goods whose consumption adversely affects health; and the existence of certain nonconsumption factors, associated with higher income occupations, which have deleterious health consequences (psychological pressure and tensions in managerial occupations, for example).[2,3]

The primary objective of this paper is to extend the research on the general question "Is income a significant factor in explaining geographic variations in mortality rates?" This is done in two ways. First, income is separated into various components to investigate whether different types of income have different effects on mortality rates. Second, the analysis is based on detailed mortality and sociodemographic data, which are aggregated into eight adult and four infant age- sex-race-specific cohorts. Thus, these data are more detailed than those used by many earlier studies.

Prior Research

Much of the earlier work done on variations in mortality rates generally can be classified into two groups: a) epidemiologically oriented studies, basically concerned with identifying differences in mortality rates for population groups classified by different socioeconomic criteria and b) those studies concerned with explaining observed differences in mortality rates by applying more formal theoretical models and multivariate analytic techniques. Among the former are works by Altenderfer, Antonovsky, and Lerner and Stutz.[4,5,6] (See Kitagawa and Hauser for a thorough review of this extensive literature.[7]) These researchers investigated various geographic units, time periods, and mortality rates. All found an inverse relationship between socioeconomic status and mortality. This basic finding is summarized by Antonovsky's observation that "Despite the multiplicity of methods in the 30-odd studies cited, and despite variegated populations surveyed, the inescapable conclusion is that class influences one's chances of staying alive. Almost without exception, the evidence shows that classes differ on mortality rates."[5] Two things to be noted about these studies, however, are the aggregate nature of the mortality indices used and the crude and unrefined measures of income employed.

Turning to the second group of studies, Auster et al., following earlier work by Fuchs, proposed a health production model to examine the relationship between age-adjusted death rates, measures of medical care consumption and a number of environmental variables, including income.‡ [2,8] They found that income is related positively to death rates and that this relation remains the same after dropping education from their analysis. They speculated that above some point, rising income may be associated with occupational conditions (such as stress, lack of exercise) and behavior changes (such as in diets, increased auto travel) that have adverse effects on mortality.

---

‡ The authors in this study were more concerned with estimating the impact of medical care on mortality rates than estimating the effects of income.

Although this study made an important contribution in this area, it has at least two shortcomings. First, as the authors themselves point out, it would have been preferable to estimate separate age- sex- race-specific production functions, since earlier work suggests that there are important differences in the factors affecting specific mortality rates. A second problem is the use of the state as the unit of analysis, since extensive intrastate heterogeneity may be masked by the use of state averages.

Silver extended the work of Auster et al., by disaggregating the mortality data by race and sex and estimating the relationships for both states and SMSAs.[3] In general, the results did not tend to support the Auster et al. findings of a positive relationship between household income and mortality (although the state-level estimates for white males were positive and significant). However, as the author cautions, these findings are derived from regressions which exclude measures of schooling, physicians per capita and fertility on the grounds of multicollinearity. When income was divided into its labor and nonlabor components, the effect of labor income tended to be positive (or less negative) while the coefficient of nonlabor income was consistently negative. The positive effect of labor income was strongest for white males, which suggests that for this subset of the population, the factors associated with increasing income may have deleterious side effects.

Grossman, in another study, concluded "Perhaps the most striking finding in this case is that health has a negative income elasticity. One explanation of this result is that the income elasticities of the detrimental inputs in the health production function exceed those of the beneficial inputs, . . ."[9] Orcutt et al. estimated variations in death rates across county groups.[10] Their analysis removed the effects of person-specific factors (age, sex, race, etc.) and attributed any residual variations to other factors such as income. They concluded that higher incomes are correlated with lower death rates. Finally, Margaret Reid examined the conflicting evidence on the relationship between death rates and income.[11] Her analysis, which covered white populations of states for 1970, found little or no evidence to support the view that death rates tend to be positively correlated with income. Although the author offered no concrete explanation as to why her income and education coefficients were not as strongly negative as those of earlier studies, she did suggest that use of age- sex-race-specific death rates as opposed to age-adjusted death rates might lead to a better understanding of the income-mortality paradox.

Empirical Methods

This section describes the empirical model that underlies the analysis. Specific issues discussed are the choice of a unit of analysis, the choice of mortality as the dependent variable, data sources, econometric issues (simultaneous equation estimation), specification of the estimating equation and variable definitions.

A. Unit of Analysis and Data

The geographic unit of analysis for this study is the county group, as defined by the Bureau of the Census. The entire country is divided into areas and subareas, each of which has a population of at least 250,000 people.[2] § (County groups may cross state boundaries.) There were 407 identified areas and subareas in 1970.

---

§ Each area is an extension of the SMSA concept—an urban center and surrounding counties tied to that center by trade and commuting patterns. In rural areas which do not contain an SMSA, the urban center is defined as a city with 20,000 to 50,000 population. Identified subareas within each county group are SMSAs or components of SMSAs with populations greater than 250,000 people.

Use of the county group as the unit of analysis offers more observations than use of either states or SMSAs. In addition, there should be less intra-area variation in the data in a county group than in a state. At the same time, county groups provide information on rural areas, since the presence of a central city of more than 50,000 people is not required to construct a county group. Compared to individual counties, the county group contains a much larger population for analysis and is not as likely to be affected by the border-crossing problem, particularly since the county group can cross state boundaries.

Within each county group, the population was divided into 12 cohorts consisting of three age groupings, 0 to 1, 45 to 64, and 65 and older, for both sexes by race (white and black). (These cohorts account for about 90 per cent of all deaths.) The main reason for disaggregating the population in this way is to control for the effects of variations in biological and sociocultural characteristics that may be associated with age, sex and race but are difficult to measure directly. The most obvious factor is natural changes in health stock usually associated with aging. Similarly, attitudes about health and medical care, the effects of racial discrimination, and inherited physical traits may be other factors that are difficult to observe but can be controlled partially by investigating age- sex- race-specific population cohorts.

Sociodemographic characteristics of each cohort were constructed by aggregating individual data from a 2 per cent sample of the 1970 census to the county group level. In order to increase the reliability of the data, a minimum of 2,000 people in a particular cohort was required for inclusion in the analysis. (For the infant analysis, a minimum of 20 deaths in each sex-race category was required.) Finally, the dependent variable is the mortality rate from all causes of death, disaggregated by age, sex and race. These rates are based on deaths over the period 1968 through the first half of 1972 in order to lessen random fluctuations in death rates.[12]

B. Empirical Specification

Preliminary analyses, reported in an earlier article, suggested the use of the log-linear functional form.[13] Thus, the dependent variable in the specifications reported below is the log of mortality from all causes. The independent variables attempt to capture the effects of initial health status, environmental conditions and medical care consumption. Table 1 lists these variables and reports their mean values for each adult cohort. Table 2 reports the mean values of the infant cohort variables, which describe characteristics of births and of women of child-bearing age (14 to 44).

The basic equations were estimated with several alternative definitions of income: Total Family Income (TFI)—the sum of the dollar amounts of wages, salary income, net nonfarm and farm self employment and other income of all household members 14 years old and over; Earned Income (EI)—the sum of wages, salaries, commissions, bonuses or tips from all jobs and self-employment, professional and farm income before deductions (excluding military bonuses, reimbursements and pay in kind); Unearned Income (UI)—total family income less earned income (TFI-EI); Social Security Income (SSI)—U.S. government payments to retired persons, to dependents of deceased insured workers or to disabled workers (excluding Medicare reimbursements); Welfare Income (WI)—amounts received from federal, state and local public programs such as AFDC, old age assistance, general assistance, and aid to the blind or totally disabled; Transfer Income (TI)—the sum of SSI and WI; and Other Unearned Income (OUI)—interest, dividends, veterans' payments, retirement pensions from private employers, unions and other governmental agencies, rental

TABLE 1. Mean Values of Regression Variables, Adult Cohorts

	WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+
LALLR—Log deaths from all causes	2.63	4.33	3.13	4.28	1.88	3.83	2.63	3.94
AHG—Average highest grade	12.61	10.48	9.72	7.21	12.72	10.98	10.46	8.11
MTJA—Mean temperature jan	34.94	34.94	40.08	43.03	39.94	34.94	40.32	41.83
PWS—% Public water system	36.25	36.43	40.00	36.95	37.43	39.16	40.80	40.17
EXALCIR*—Excess alcohol and cirrhosis deaths	0.0001	−0.0057	0.0026	−0.248	0.0003	−0.0057	−0.0043	0.0054
AVA—Average age	54.28	74.64	54.21	73.52	54.33	74.80	54.20	74.07
USB—Per cent U.S. born	95.59	86.93	98.82	98.51	95.33	89.18	98.94	98.94
MIGR—Per cent migrants	2.97	2.18	1.79	1.23	2.89	2.31	1.84	1.23
EXFLU*—Excess flu deaths	0.0020	0.0017	0.0080	0.0003	0.0006	−0.0002	0.0072	0.0014
NILF—Per cent not in labor force	N.A.	N.A.	N.A.	N.A.	36.17	73.76	27.03	66.12
KIDS—Average no. children born	N.A.	N.A.	N.A.	N.A.	2.51	2.81	3.07	3.13
DISABLH†—Per cent work-limiting disability	10.13	N.A.	12.60	N.A.	8.41	N.A.	13.24	N.A.
LMCDREXP‡—Log medicare expenditures per enrollee	5.68	5.68	5.69	5.65	5.68	5.68	5.68	5.64
MAR—Per cent married	88.09	72.41	73.59	59.81	75.72	35.27	56.55	27.22
CIGS§—Cigarettes sold per person (00s)	23.78	23.78	23.96	23.72	23.78	23.78	23.85	23.58
VET—Per cent veteran	25.78	13.16	20.55	10.28	N.A.	N.A.	N.A.	N.A.
WORK68—Per cent last worked after 1968	10.89	17.20	13.25	15.75	N.A.	N.A.	N.A.	N.A.

* Defined as the difference between actual and predicted deaths.

† Predicated value from a first-stage regression.

‡ Values for white cohorts are predicted from the first-stage regression of Medicare expenditures per enrollee with all of the independent variables in the model.

§ Defined for each cohort's state.

N.A. = Not applicable/available.

TABLE 2. Mean Values of Regression Variables, Infant Cohorts*

	WF 14–44	BF 14–44
LMIR—Log male infant death rate	2.97	3.55
LFIR—Log female infant death rate	2.70	3.35
AHG—average highest grade	13.52	12.20
MAR—Per cent married	87.35	57.14
NILF—Per cent not in labor force	27.18	19.44
USB—Per cent U.S. born	95.84	98.78
KIDS—Average no. of children born	2.83	3.29
SPAN—Per cent Spanish	4.09	1.32
CIGS†—Cigarettes sold per person (00s)	23.76	23.63
LMCDREXP—Log medicare expenditures per enrollee	5.69	5.68
LEMIR--Log expected male infant death	3.23	3.91
LEFIR--Log expected female infant death	2.96	3.68
PMBHFM‡—Per cent male births high risk mothers	6.70	7.90
PFBHRM‡—Per cent female birth high risk mothers	6.80	8.00
PMBIH—Per cent male births in hospitals	99.70	95.70
PFBIH—Per cent female births in hospitals	99.70	95.50

* AHG, MAR, NILF, USB, ANKB, and SPAN are based on data from the 1970 Census for women of child-bearing age (14 to 44). Data on births and deaths are from vital statistics records.

† Defined for the state.

‡ Mothers younger than 15 and older than 34.

income, workman's compensation, alimony, private welfare payments and contributions from persons who are not members of households.

Table 3 reports the means of these variables for each cohort. As is well known, whites have higher family incomes than blacks in each age-sex group, by factors ranging from 60 to 85 per cent. White-black differences are much smaller for transfer incomes from public sources (Social Security payments and welfare income). In fact, blacks have higher levels of transfer income than whites except in the two elderly age groups. Differences are much greater, on the other hand, for transfer and unearned income from private sources (OUI), with whites having from two to six times more income from these sources than blacks.

All equations were estimated by ordinary least squares (OLS). In addition, the adult-cohort equations also were estimated by a simultaneous equation method, instrumental variable (IV) estimation. The latter approach is justified on the grounds that for adults, both medical care and income may be simultaneously determined with health. Thus, each of the income variables and the medical care use variable were replaced by predicted values generated by regressing income and medical care against all of the exogenous variables specified by a larger model of health production.[13]

Preliminary results indicated that the OLS and IV estimates of the coefficients of the income variables were very similar. This suggests that, at least for these data, simultaneous determination of mortality rates and the various income measures may not be a problem. The IV estimates of the coefficients of the medical care variable, however, did differ from the OLS estimates for the white cohorts, but not for the black cohorts. Based on these preliminary results, therefore, the income coefficients reported below for adults are based on OLS estimation of the black cohorts' mortality regressions and IV estimation of the white cohorts' equations (with medical care use replaced by its predicted value).

As noted previously, one of the explanations offered for the income–mortality

TABLE 3. Mean Values of Income Variables, by Cohort

	WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+	WF, 14–44	BF, 14–44
TFI—Total family income	$11,791	$6,570	$7,272	$3,941	$10,376	$5,863	$6,226	$3,811	$11,691	$6,821
EI—Earned income	8,242	1,638	4,377	857	2,060	405	1,559	303	1,971	2,012
UI—Unearned income	3,463	3,808	2,733	2,225	8,229	4,732	4,466	2,842	9,720	4,609
SSI—Social security income	83	1,089	110	730	83	666	101	468	18	35
WI—Welfare income	3	35	52	129	4	59	99	198	16	130
TI—Transfer income	86	1,125	162	859	87	725	200	666	34	165
OUI—Other unearned income	603	1,327	143	268	265	673	627	113	104	39

Definitions: *Total family income*—The sum of the dollar amounts of wages, salary income, net non-farm and farm self employment, and other income of *all* household members 14 years old and over. *Earned income*—Includes wages, salaries, commissions, bonuses or tips from all jobs and self employment, professional, and farm income before deductions (excluding military bonuses, reimbursement, and pay in kind). *Unearned income*—Total family income less earned income. *Social Security income*—Includes U.S. government payments to retired persons, to dependents of deceased insured workers, or to disabled workers (excluding Medicare Reimbursements). *Welfare income*—Includes amounts received from federal, state and local public programs such as AFDC, old age assistance, general assistance, and aid to the blind or totally disabled. *Transfer income*—The sum of S.S.I. and W.I. *Other unearned income*—Includes interest, dividends, veterans payments, retirement pensions from private employers, unions and other governmental agencies, rental income, workman's compensation, alimony, private welfare payments, and contributions from persons not members of households.

paradox is that the underlying relationship is nonlinear, i.e., mortality rates fall as income rises, but in successively smaller decrements. Eventually, on the "flat" portion of the curve, further increases in income will have no impact on mortality rates. This hypothesis can be tested by adding an income-squared variable to the model. Its expected sign is positive. Preliminary tests of this nonlinearity hypothesis suggested that the functional form reported below is superior.[e] This result also suggests that differences in the variance of the income distributions of the cohorts within a county group can be ignored. If the income-mortality relationship is approximately linear over the relevant range, then the experiences of persons in the upper and lower tails of the income distribution will offset each other.[f]

Finally, several earlier studies have emphasized the difficulty of untangling the effects of income on mortality rates from those of education.[2,3,9] This problem is addressed by estimating each equation both with and without the education variable in order to gauge the strength of the interrelationship between these factors. In all, 14 specifications are estimated. The last seven are essentially the same as the first seven, except they exclude education. The seven main specifications differ with respect to the definition of income.

Results

The coefficients of the adult cohorts' income variables (from the equations with education included) are reported in Table 4. Although the underlying equations explain exceedingly high proportions of the variance in mortality from all causes (R^2s vary around 0.9), these coefficients do not provide an unambiguous resolution of the income–mortality issue. The frequency of negative coefficients appears to tilt the balance in favor of an inverse relationship, but a number of the coefficients are not statistically significant. In Specification 1, for example, the coefficients of TFI, total family income, are negative for five of the cohorts, but statistically significant only for the two middle-aged male cohorts.

When total family income is decomposed into its earned (EI) and unearned (UI) components, a different pattern emerges. Earned income coefficients are negative and significant for three of four middle-aged cohorts and positive for three of four elderly cohorts. For unearned income however, the signs of the coefficients are reversed between middle-aged and elderly cohorts. In general, the pattern is that EI has a negative effect for the younger cohorts and positive effect for elderly cohorts. This should not be too surprising, however, since this variable partially captures the influence of initial health, which affects both labor force participation and mortality rates. People in good health and of greater economic means generally have higher earned incomes while of working age, and higher unearned income (and thus less need to work) while of retirement age.

Specifications 3 and 4 estimate the impact of earned and unearned income separately. The results suggest no significant qualitative or quantitative differences from Specification 2 on the impact of both EI and UI on mortality rates either through joint or independent estimation. The coefficients of both earned and unearned income are essentially identical to those estimated by Specification 2.

---

[e] In seven of the eight adult cohorts, both the linear and squared terms entered with implausible signs, positive and negative, respectively. This pattern implies that mortality rates initially increase as income increases, but eventually decline in successively larger decrements. A double-log specification also was tested, but its implications were essentially the same as those reported below.

[f] Another factor which lessens the need to account explicitly for the variance of income is stratification of the population into age- sex- race-specific cohorts. As Table 3 indicates, intercohort variation is larger than intracohort variation. The mean coefficient of variation across cohorts is 0.31. When the cohorts are pooled, the coefficient of variation jumps to 0.45.

The remaining specifications focus on different components of unearned income. As noted above, UI includes the earnings of other family members as well as rental and capital income, and various transfer payments from public and private sources. Looking first at public transfer programs, Specification 5 examines the impact of transfer income. In general, the results are ambiguous. Although five of the coefficients are negative, only two are statistically significant. However, two of the three positive coefficients are also statistically significant. These results do not lend themselves to meaningful interpretation.

Decomposing public transfer payments into Social Security payments and welfare payments helps somewhat to clarify the ambiguity of the TI coefficients. In particular, seven of the eight coefficients of welfare income are negative, with four being statistically significant. Conversely, six of the SSI coefficients have positive signs, with three being statistically significant. Since SSI includes payments to deceased insurees' spouses, who generally have higher mortality rates than either married or single people,[14] and to the disabled, it may be that this variable is strongly associated with poor health status. The inability of the instrumental variable estimates of Specifications 5 and 6 to detect this relationship may have been the result of inadequate instrumental variables for TI, SSI and WI, rather than the absence of any relationship between health and these variables.

Finally, the last specification focuses on unearned income from capital sources and from private pension plans. The qualitative results for this variable are very similar to those for pure public transfer payments, WI. The coefficients are negative for seven of the eight cohorts and five of these also are statistically significant. The strength of this finding may be due to the fact that this

TABLE 4. O.L.S. Income Coefficients, Adult Cohorts (Education Included)

Income Specification		WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+
1.	TFI	−0.0006*	−0.0003	−0.0016†	0.0003	−0.0003	NS	−0.0003	−0.0003
2.	UI	0.0006	−0.0010‡	0.0006	−0.0009	0.0001	0.0005	0.0001	−0.0001
	EI	−0.0010†	0.0007	−0.0033†	0.0032*	−0.0032*	−0.0095†	0.0013	0.0007
3.	UI	0.0003	−0.0007‡	−0.0002	−0.0005	−0.0002	0.0001	0.0002	−0.0001
4.	EI	−0.0010†	0.0002	−0.0032†	0.0029*	−0.0030*	−0.0084†	0.0014	0.0007
5.	TI	−0.0106‡	−0.0021	0.0060	−0.0039	0.0117‡	0.0195†	−0.0031	−0.0118*
6.	SSI	−0.0073	0.0009	0.0143‡	0.0001	0.0110‡	0.0283†	0.0009	−0.0009
	WI	−0.0272‡	−0.0247†	−0.0087	−0.0223†	0.0158	−0.0118	−0.0058	−0.0023†
7.	OUI	−0.0028‡	0.0037†	−0.0040	0.0041‡	−0.0148†	−0.0052†	−0.0216*	−0.0200†

NS, No significant digits.

* Significant at the 95 per cent confidence level.

† Significant at the 99 per cent confidence level.

‡ Significant at the 90 per cent confidence level.

form of unearned income is least likely to be related to health status. As such, it may be the income component that is most closely correlated with wealth and permanent income.

The results for the infant cohorts (from the specification including education) are presented in Table 5. In general, the evidence suggests a fairly consistent negative relationship between mortality and income for all seven specifications. Coefficients are similar in magnitude for female and white male infants, but generally smaller in magnitude for black male infants. Both mothers' earned income and unearned income from all sources, which consists primarily of spouses' and other family members' earnings, have uniformly negative signs and are statistically significant in all but one case. The impact of public transfer income, which is a very small component of total family income on the average, is ambiguous. Finally, private transfer income, which makes up less than 1 per cent of total income, has negative coefficients for three of the four cohorts.

Education and income are highly correlated, as is well known. Thus, including an education variable in the regression equations increases the unreliability of the estimated income coefficients and makes their interpretation more ambiguous. Analysis of the simple correlation coefficients between TFI and education provides evidence to support this hypothesis. The correlation coefficients vary from 0.184 (elderly white males) to 0.766 (middle-aged white males) for the adult cohorts and 0.561 (blacks) to 0.770 (whites) for the infant cohorts (women of childbearing age). However, the apparent presence of a relatively high degree of multicollinearity is not sufficient reason to follow Silver and exclude education from our equations.[3] First, there is ample evidence in the literature to suggest the relevance of education as an important factor in explaining variations in mortality rates. Second, the exclusion of a relevant variable such as education may lead to serious error: not only may the estimated coefficients be biased, but inferences based on them also will be inaccurate. As Johnston suggests, "Data and degrees of freedom permitting, one should err on the side of including variables in the regression analysis rather than excluding them."[15]

In order to gauge the impact of multicollinearity on the results, we also report the income coefficients from equations that exclude the education variable. These results are presented in Tables 6 and 7. In

TABLE 5. O.L.S. Income Coefficients, Infant Cohorts (Education Included)

Income Specification		Males		Females	
		Blacks	Whites	Blacks	Whites
1.	TFI	−0.0007	−0.0016*	−0.0023†	−0.0019*
2.	UI	0.0012	−0.0011*	−0.0014	−0.0017*
	EI	−0.0063‡	−0.0085*	−0.0048	−0.0043‡
3.	UI	−0.0001	−0.0014*	−0.0024‡	−0.0019*
4.	EI	−0.0051‡	−0.0106*	−0.0063‡	−0.0076*
5.	TI	0.0002	0.0110	−0.0036	0.0015
6.	SSI	−0.0243	0.0197‡	−0.0406†	0.0068
	WI	0.0114	0.0008	0.0128	−0.0047
7.	OUI	−0.0093	−0.0136‡	0.0142	−0.0112‡

* Significant at the 99 per cent confidence level.
† Significant at the 95 per cent confidence level.
‡ Significant at the 90 per cent confidence level.

general, the signs of the coefficients suggest an inverse relationship between income and mortality. As expected, most of the coefficients are biased negatively.# Only one coefficient (elderly black males) appears to have a positive bias. For total family income (TFI) the average percentage increase in absolute value for the negative coefficients is about 50 per cent.

A similar analysis of the effect of income on the education coefficients can be conducted by estimating regressions that include and exclude the income variables. The result of such an analysis for all cohorts is presented in Table 8, where the income variable in question is total family income. As can be gleaned from the table, education coefficients, with total family income excluded, are, in general, negatively biased and larger in absolute value than those with family income included. Nine of 12 coefficients increased in absolute value, on the average by about 60 per cent, with the maximum and minimum changes occurring for middle-aged white females and elderly black males, respectively. In general, education has a consistently negative association with mortality rates.

Conclusions

Several research and policy issues were raised at the outset of this paper. Is income a significant factor in explaining geographic variations in adult mortality rates? What are the differential effects (if any) of the various components of income? What is the effect of "pure" unearned income on mortality rates?

These questions were addressed through estimation of a health production function that used an inverse indicator, the mortality rate from all causes, as the dependent variable. The primary empirical

TABLE 6. O.L.S. Income Coefficients, Adult Cohorts (Education Excluded)

Income Specification		WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+
1.	TFI	−0.0011*	−0.0003	−0.0016*	−0.0004	−0.0002	−0.0004†	NS	−0.0005
2.	UI	0.0005	−0.0006‡	0.0009	−0.0009	0.0001	−0.0003	0.0002	−0.0005
	EI	−0.0017*	0.0007	−0.0034*	0.0007	−0.0026†	−0.0034†	−0.0004	−0.0014
3.	UI	−0.0002	−0.0005†	NS	−0.0008	−0.0001	−0.0004†	0.0001	−0.0005
4.	EI	−0.0017*	0.0002	−0.0032*	0.0004	−0.0023†	−0.0039†	−0.0003	−0.0016
5.	TI	0.0029	−0.0033†	0.0023	0.0032	0.0128‡	0.0126*	−0.0057	−0.0017
6.	SSI	0.0075	−0.0029	0.0117†	0.0059†	0.0121†	0.0203*	−0.0020	−0.0014
	WI	−0.0196	−0.0095†	−0.0138†	−0.0095	0.0162	0.0030	−0.0081	−0.0020
7.	OUI	−0.0051*	−9.0028*	−0.0059†	0.0048†	−0.0131*	−0.0044*	−0.0225‡	−0.0083

NS, No significant digits.
* Significant at the 99 per cent confidence level.
† Significant at the 90 per cent confidence level.
‡ Significant at the 95 per cent confidence level.

On theoretical grounds, if education is hypothesized to be positively related to income but negatively related to mortality, then omitting education from the regression equation should negatively bias the income coefficients.

TABLE 7. O.L.S. Income Coefficients, Infant Cohorts (Education Excluded)

Income Specification		Males		Females	
		Blacks	Whites	Blacks	Whites
1.	TFI	−0.0011	−0.0021*	−0.0029†	−0.0026*
2.	UI	0.0012	−0.0016*	−0.0015	−0.0024*
	EI	−0.0067†	−0.0094*	−0.0060‡	−0.0053‡
3.	UI	−0.0007	−0.0020*	−0.0032†	−0.0026*
4.	EI	−0.0053†	−0.0133*	−0.0078*	−0.0111*
5.	TI	−0.0035	0.0146‡	−0.0099	0.0061
6.	SSI	−0.0264	0.0128	−0.0443‡	−0.0008
	WI	0.0066	0.0167	0.0050	0.0142
7.	OUI	−0.0114	−0.0290*	0.0091	−0.0315*

Income and education are based on data for women of child-bearing age.

* Significant at the 99 per cent confidence level.

† Significant at the 95 per cent confidence level.

‡ Significant at the 90 per cent confidence level.

extension offered by the work reported here was estimation of these functions with seven more refined measures of income entering the specifications. Separate relationships were estimated for 12 population cohorts consisting of three age groups, 0 to 1, 45 to 64, and 65 and older, for both sexes by race (white and black).

Based on the results reported above, the answer to the first question would seem to be a qualified yes. Analysis of all cohorts and specifications reveals a negative relationship in general. Although a number of the estimated coefficients were statistically insignificant, particularly for adults, examination of coefficients from variants

TABLE 8. Coefficients of Education, All Cohorts*

Cohorts	Pearson Correlation Coefficient†	Income Included	Income Excluded
Adults			
WM, 45–64	0.766	−0.0205‡	−0.0239‡
WM, 65+	0.533	−0.0034	−0.0122‡
BM, 45–64	0.702	−0.0132§	−0.0203‡
BM, 65+	0.533	−0.0104§	−0.0099
WF, 45–64	0.631	−0.0165‖	−0.0150‡
WF, 65+	0.184	−0.0203‡	−0.0233
BF, 45–64	0.600	−0.0382‡	−0.0418‡
BF, 65+	0.431	−0.0126	−0.0099
Male Infants			
Whites	0.770	−0.0522‡	−0.0754‡
Blacks	0.561	−0.0180	−0.0252
Female Infants			
Whites	0.770	−0.0651‡	−0.0945
Blacks	0.560	−0.0242	−0.0479

* Income = total family income.

† Income and education.

‡ Significant at the 99 per cent confidence level.

§ Significant at the 90 per cent confidence level.

‖ Significant at the 95 per cent confidence level.

that excluded education suggested that multicollinearity reduces the reliability of the estimates. At the same time, however, it is clear that excluding education results in the income coefficients being negatively biased.

With respect to the differential effects of various components of income, our analysis shows that earned income has a negative impact on middle-aged cohorts, but a positive impact on elderly cohorts. The reverse seems to hold true for unearned income. This finding, with respect to the effects of earned and unearned income, is inconsistent with Silver's study.[3] A plausible explanation of this inconsistency is that Silver's results probably are dominated by results for the elderly, since changes in mortality for these cohorts have a major impact on age-adjusted mortality. However, the results of the pure income transfer variables, WI and OUI, tend to reinforce and clarify Silver's conclusion that higher unearned income contributes to lower mortality rates. Thus, the somewhat ambiguous results for total family income can be understood better when income is broken down into earned and pure transfer components. The impact of the former is associated with age, while the latter generally has a negative association with mortality rates for all cohorts.

Another indicator of the effect of income on mortality is its elasticity, that is, the percentage change in mortality for a small change in income. Table 9 reports elasticities (evaluated at the means) for TFI and OUI, the two variables that most consistently had negative coefficients. The evidence gathered from this table is that income has a definite negative impact on mortality, but the strength of the relationship is rather weak. For both income variables considered, a 1 per cent change in income reduces mortality by about 0.05 per cent on the average. The effect is largest for the infant cohorts, with "family income elasticities" ranging from −0.047 to −0.222. On the other hand, the elasticity of OUI is larger than that for TFI for five of eight adult cohorts.

TABLE 9. Income Elasticities, All Cohorts

Cohorts	TFI*	OUI*
Adults		
WM, 45–64	−0.071	−0.017
WM, 65+	−0.020	−0.049
BM, 45–64	−0.116	−0.005
BM, 65+	0.012	0.011
WF, 45–64	−0.031	−0.039
WF, 65+	0.000	−0.035
BF, 45–64	0.019	−0.013
BF, 65+	−0.011	−0.023
Male Infants		
Whites	−0.187	−0.014
Blacks	−0.047	−0.004
Female Infants		
Whites	−0.222	−0.012
Blacks	−0.157	0.006

* Based on coefficients of Total Family Income (TFI) and Other Unearned Income (OUI) from Tables 5 and 6, and mean values of TFI and OUI from Table 3.

From a policy viewpoint, these results indicate that income transfer policies aimed at reducing mortality rates generally will contribute to lower mortality, but the magnitude of the effect will not be large. Since reducing mortality rates generally is not the primary motive for income transfers, this conclusion should not be interpreted as a justification for cutting back income transfer programs. Rather, this result reinforces the argument that transfer programs have payoffs in addition to the financing of current consumption.

References

1. American Medical Association. Report of the commission on the cost of medical care, vol 3. Chicago: American Medical Association, 1963-64.

2. Auster R, Levenson I, Sarachek D. The production of health, an exploratory study. J Hum Resour 1969;4:412.

3. Silver M. An economic analysis of spatial variations in mortality rates by age and sex. In: Fuchs VR, ed. Essays in the economics of health and medical care. New York: National Bureau of Economic Research, 1972.

4. Altenderfer ME. Relationship between per capita income and mortality in the cities of 100,000 or more population. Public Health Rep 1947;62:1681.

5. Antonovsky A. Social class, life expectancy and overall mortality. Milbank Mem Fund Q 1967;48:31.

6. Lerner M, Stutz RN. Mortality by socioeconomic status, 1959-61 and 1969-71. Md State Med J 1978.

7. Kitigawa EM, Hauser PM. Differential mortality in the United States. Cambridge: Harvard University Press, 1973.

8. Fuchs VR. The contribution of health services to the American economy. Milbank Mem Fund Q 1966;44:65.

9. Grossman M. The demand for health: a theoretical and empirical investigation. New York: National Bureau of Economic Research, 1972.

10. Orcutt GH, Franklin SR, Mendelsohn R, Smith JR. Does your probability of death depend on your environment? A microanalytic study. American Economic Review 1977;67:260.

11. Reid M. An exploration of conflicting evidence of death rates and income. Unpublished data, Department of Economics, University of Chicago, 1976.

12. U.S. Department of Health and Human Services, National Center for Health Statistics. Geographic patterns in the risk of dying and associated factors ages 35-74 years, United States, 1968-72. Hyattsville, Md.: USDHHS, NCHS, 1980.

13. Hadley J. Assessing the adequacy of health manpower supply. Urban Institute Working Paper 1225-5. Washington, D.C., November 1980.

14. Sheps MC. Marriage and mortality. Am J Public Health 1961;51:547.

15. Johnston J. Econometric methods. New York: McGraw-Hill, 1972.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Income and Race Differences in Children's Health in the Mid-1960s

LINDA N. EDWARDS, PH.D.,* AND MICHAEL GROSSMAN, PH.D.†

This article explores income and race differences in eight measures of the health of children ages 6 through 11 as assessed in the early 1960s. It is shown that both income and race differences in health are much less pronounced than they are in infant mortality and birth weight data. Significant differences are found in the health status of black and white children and of children from high- and low-income families, but these are primarily differences with respect to parent-reported (rather than physician-reported) health criteria and they by no means overwhelmingly favor the white or high-income children. These findings underscore the importance of treating children's health status as multidimensional. In addition, these findings will serve as a bench mark for studies of children's health using data for a more recent period.

RECENT REVIEWS of children's health in the United States have stressed the differences between black and white children and between children in high- and low-income families.[1,2] These studies cite mortality rates for both infants and older children that are over 50 per cent higher for blacks than for whites. They also point out that low birth weight (a good negative indicator of whether an infant will survive his first year and of his successful future development) is more prevalent among black and poor families.

The income and race comparisons in children's health status cited above, as well as those cited elsewhere, rely primarily on measures that relate to the first year of the child's life. This results largely from the unavailability of comprehensive measures of morbidity for older children. Publication of data from Cycles II and III of the Health Examination Survey, however, makes it possible to study explicitly the health of older children. In this paper we use data from Cycle II, which covers children aged 6 through 11 years in the 1963-65 period, to explore income and race differences in eight measures of children's health.

The Cycle II children were examined in the mid-1960s. Nevertheless, their records are still an important resource for analysis. These children comprise the only large representative sample of U.S. children taken from the recent past for which such a rich array of health and family background data are available. Undoubtedly, there have been changes in the health status of children in this age-group during the intervening 15 years, especially as a result of goverament health programs targeted at the poor, so that the relative status of black and poor children may have changed. But

---

* Department of Economics, Queens College and Graduate School, of the City University of New York and National Bureau of Economic Research, New York, New York.

† Department of Economics Graduate School, of the City University of New York and National Bureau of Economic Research, New York, New York.

This paper has not undergone the review accorded official National Bureau of Economic Research publications; in particular, it has not yet been submitted for approval by the Board of Directors.

Supported by PHS Grant No. 1 R01 HS 02917 from the National Center for Health Services Research to the National Bureau of Economic Research and a grant from the Robert Wood Johnson Foundation to the NBER.

Reprint requests: Michael Grossman, National Bureau of Economic Research, 269 Mercer Street, 8th Floor, New York, New York 10003.

0025-7079/82/0900/0915/$01.60 © J. B. Lippincott Co.

even so, analysis of the data from 1963 to 1965 will provide a benchmark against which to measure that change as future samples become available.

We report two kinds of results. First, we show that when health measures from mid-childhood are the subject of analysis, both income and race differences are much less pronounced than they are in infant mortality data. Indeed, for one of our most comprehensive health measures—the examining doctor's assessment of whether or not the child has any "significant acquired abnormalities"—there are no statistically significant differences either by income or by race. Second, the differences that we do find do not uniformly favor children from white or higher-income families.

Measuring Health in the Cycle II Sample

Cycle II of the Health Examination Survey, conducted by the National Center for Health Statistics (NCHS), is an exceptional source of information about a national sample of 7,119 noninstitutionalized children aged 6 to 11 years in the 1963-65 period. The data comprise medical and developmental histories of each child provided by the parent, information on family socioeconomic characteristics, birth certificate information and a school report with data on school performance and classroom behavior provided by teachers or other school officials. Most important, there are objective measures of health from physical examinations administered by the Public Health Service. Together, the data provide an unusually detailed picture of the health of children in this cohort.

A full description of the sample, the sampling technique and the data collection is presented in NCHS.[3] An exceptional strength of this sample is its high response rate of 96 per cent with nonwhites having a slightly higher response rate than did whites (0.982 versus 0.956). (See NCHS,[3] Table 4.) The one deficiency of the sample, from the point of view of studying children's health, is the exclusion of children in institutions. To the extent that these children are more likely to have serious and disabling physical conditions, the reported incidence of certain conditions will be lower in our sample than in the entire population of children. In addition, if the probability of the institutionalization of a child with a given condition depends on income and race, our results will incorporate unknown biases. The number of institutionalized children is small, however, at about four-tenths of a per cent of all children aged 5 through 13 years. This is the proportion of 5–13 year olds living in "group quarters" in 1970 according to the U.S. Bureau of the Census,[4] Tables 52 and 205. The corresponding percentages by race are 0.38 per cent for whites and 0.7 per cent for blacks.

Given the unusual detail and diversity of the health data in the Cycle II survey, the choice of health status measures is not dictated solely by data availability, but rather is an issue that must be faced directly. The problem of defining and measuring children's (and adults') health is very much an unresolved one, even among professionals in the area of public health.[5,6] The economist's approach is to define health as a form of human capital that determines the amount of time available for consumption and for work in the home and labor market.[7] With this type of definition, an appropriate measure of health status over some time period would be the proportion of one's potential time that is actually available for the usual consumption, maintenance and work activities. Similarly, the complementary measure of ill health would be that proportion of one's potential time lost as a consequence of imperfect functioning. Such disability may be relatively easy to measure when dealing with adults who are members of the labor force (a good approximation is days lost from work because of illness), but it is not easy to measure for other adults or for children. Therefore, in economists' studies of adults' health, the occurrence of particular physical conditions and the individual's own assessment of his health status have been used as supplementary health measures.[8]

We use the same type of restricted, morbidity-oriented, definition of children's health, focusing on the child's physical health rather than his overall well-being, and similar types of measures including disability, physical conditions, and parental assessment of health status. Even some of these measures, however, may not always be appropriate for children, because certain childhood diseases and acute conditions that prevent children from carrying out their normal activities may not reflect their health capital or "permanent" health. A useful distinction to make here is between permanent health which is one's prospect for life preservation and normal lifetime functioning, and "transitory" health, which is short-run deviations from one's normal state of health. It is the child's permanent health status that we wish to study, and we attempt to choose health measures that are indicators of that permanent health status.

Of course, there is a positive relationship between permanent and transitory health status, in the sense that a child with low health capital is more likely to contract some acute conditions and to have them for a more extended time period. For example, Birch and Gussow[9] discuss how a child's nutrition (which is clearly a determinant of permanent health status) and his likelihood of sustaining an acute condition are intimately related. Similarly, Starfield[10] reports that some acute problems may have a high likelihood of permanence in some population subgroups.

In some situations a single overall index of permanent health might be desired to help one describe parsimoniously the health status of a population, for example, or allocate public funds. Infant mortality statistics frequently are used in this way. Health, however, is clearly a multidimensional concept. Consequently, we use a set of health measures rather than a single index. Analysis of a set of component measures also avoids the essentially arbitrary weighting of the various dimensions of health implied by a health index.[‡] Finally, such analysis allows for the possibility that the various aspects of children's health are differentially related to family income and race.

The actual choice of components of children's health status to be examined has been guided by the child health literature,[11–19] as well as by discussions with public health specialists and pediatricians. In addition, we have tried to take advantage of the diversity of the data by including measures derived not only from the physician's examination but also from the parent's report and the school report. The latter two types of measures, of course, are subjective, in that they depend on the perceptions of the parent or teacher. The final set of measures we study, listed and described below, will not satisfy everyone. But some exclusions result from unavailability (anemia and corrected distance vision, for example) or extremely low prevalence (poor hearing and injuries due to accidents), rather than deliberate choice.

1. The presence of one or more "significant acquired abnormalities" on physical examination of the child, represented by the dichotomous variable ABNORMALITY. These abnormalities include heart disease; neurological, muscular or joint conditions; and other major diseases.[§]

2. The child's visual acuity, represented by the dichotomous variable VISION. VI-

---

[‡] In earlier work some attempts were made to condense the health information with the use of principal component analysis. The analysis yielded almost as many equally weighted components as there were initial health measures.

[§] In defining ABNORMALITY, we exclude abnormalities resulting from accidents or injuries, because these may be more likely to reflect transitory rather than permanent health variations. The proportion of children in the full sample that had abnormalities due to accidents or injuries (but no other type of abnormality) is 2 per cent. Note that congenital abnormalities are not studied, because they generally relate to events that occur prior to birth (the prenatal period is not investigated in this paper).

SION indicates whether the child has abnormal distance vision. All children were examined only without their eyeglasses; their uncorrected binocular distance vision was defined as abnormal if it was worse than 20/30.[20] It also would have been desirable to investigate corrected distance vision, but the Cycle II data do not include the information necessary to do so.

3. The child's blood pressure, represented by BLOODP. BLOODP is a dichotomous variable which indicates if the child's diastolic blood pressure exceeds the 95th percentile for his or her age and sex.

4. The child's periodontal index, represented by PERI. The periodontal index is described as follows:[21] "Every tooth in the mouth . . . is scored according to the presence or absence of manifest signs of peridontal disease. When a portion of the free gingiva is inflamed, a score of 1 is recorded. When completely circumscribed by inflammation, teeth are scored 2. Teeth with frank peridontal pockets are scored 6 when their masticatory function is unimpaired and 8 when it is impaired. The arithemetic average of all scores is the individual's [periodontal index], which ranges from a low of 0.0 (no inflammation or periodontal pockets) to a high of 8.0 (all teeth with pockets and impaired function)."[21] PERI is a good overall indicator of oral health as well as a correlate of nutrition.[22] Since the periodontal index is known to differ systematically by age and sex, PERI is standardized by age and sex (PERI is measured as the difference between the child's actual periodontal index and the average value of that index for his or her age-sex group divided by the standard deviation of the index for that age-sex group). Higher values of PERI denote poorer values of oral health.

The periodontal index, however, suffers from the defect that it is subject to intrarater and inter-rater variability. We have experimented with a somewhat more objective measure of oral health, the number of decayed permanent and primary teeth adjusted for age and sex, and have obtained results similar to those for the periodontal index. We chose to report on the periodontal index rather than on the prevalence of decay because we judged that the former is more likely to reflect underlying nutritional and health habits and less likely to simply indicate the receipt of recent dental care.

5. The parent's assessment of the child's overall current health, represented by ASSESS. ASSESS is a dichotomous variable indicating whether the parent views the child's health as either "poor," "fair" or "good" as opposed to "very good." This particular dichotomy was chosen because there were very few children rated in the "poor" and "fair" categories (5.0 per cent) whereas a large number of parents felt it pertinent to distinguish between "good" and "very good."

It is interesting to note that although this is a subjective measure, the NCHS[23] reports that it is positively related to the physician's overall assessment of the child's health. That is, children who are given a lower health rating by one of their parents are more likely than average to be rated by the physician as having a "significant abnormality."

6. Whether or not the parent reports that the child has hay fever, asthma or "other allergies," represented by the dichotomous variable ALLERGY.

7. The parent's assessment of the child's level of tension, represented by the dichotomous variable TENSION. TENSION indicates if a child is rated by his or her parent as "high strung" or "moderately tense," as opposed to "moderately relaxed" or "unusually calm and relaxed." This dichotomy, again, was partially dictated by the data; very few of the children were rated in the "high strung" category.

8. Excessive absence from school for health reasons during the past 6 months, represented by the dichotomous variable ABSENT. This variable is taken from information provided by the child's teacher or a school administrator.

As is implied in the above definitions, all of these variables are negative correlates of good health. Note that four of the variables are taken from the physical examination (ABNORMALITY, VISION, BLOODP AND PERI), three are taken from the medical history form completed by the parent (ASSESS, ALLERGY and TENSION) and one is taken from the school form (ABSENT). Note also that variables taken from the physical examination are subject to inter-rater variability in that physicians and dentists may differ in the way they interpret the results of an examination.[24,25] This means that both objective health variables (those obtained from the physical examination) and subjective health variables (those obtained from the parent) are subject to measurement errors. But measurement errors in the subjective variables are likely to be correlated with income and race (see below), while measurement errors in the objective variables are unlikely to be correlated with race and income. This follows because white and black children and children from different income classes were examined by physicians and dentists with similar qualifications.

Race and Income Differences in Health Status in the Cycle II Sample

Table 1 presents the mean levels of these eight health measures for our entire working sample, as well as by family income and by race. The two family income classes are under $5,000 per annum and $5,000 or more per annum. This income cutoff is selected because it approximately identifies the lowest quartile of the income distribution for the Cycle II sample. For purposes of comparison, Table 1 also includes statistics on infant mortality and the prevalence of low birth weight for the same income and race classes.

The working sample used in Table 1, as well as in all subsequent calculations, includes only 4,777 of the 7,119 observations, because, first, we restricted our analysis to children who either lived with both of their parents or with their mothers only (no stepparents, foster parents, grandparents, etc.). We did so to guarantee that the family background variables used in the next section reflect the actual environment of the child.[§] We also excluded those children for whom there were missing data for any of the health measures or socioeconomic variables used in the next section.[‖] A comparison of the means of the eight health measures computed from our working sample with corresponding means for the entire Cycle II sample reveals very little difference between them in most cases (Table 2).[¶] Table 2 also reveals that the entire Cycle II sample contains slightly higher proportions of black children and children from low-income families than the working sample. These differences are not, however, large enough to suggest that the results of our analysis would be altered if the entire sample had been utilized.

Looking now at Table 1, we present not only the race and income differences in the health measures, but also the standard normal statistics required for testing whether or not observed race and income differences could have arisen by chance.[#] These "z" statistics are computed in two ways. The first way assumes that the Cycle II data were generated by a simple random sampling procedure, while the second makes an adjustment to allow for the actual multi-stage cluster sampling design used to collect the Cycle II data.

Computing sample standard errors for this type of sample is complicated and requires information about the location of each sample observation. At the time of

---

[§] Only 1 per cent of the children lived with their fathers only. The 72 children who turned 12 years old before the completion of the survey were also excluded.

[‖] There is no school form for approximately 500 children in the Cycle II data set. Since excessive absence due to illness is the only variable taken from the school form, children without the school form are eliminated from our working sample only when school absenteeism is examined.

[¶] All means and standard deviations discussed here and in the rest of the paper were computed from the unweighted data. Follow-up calculations with the weighted data altered the means only slightly (typically in the third decimal place only).

[#] For the dichotomous health measure the appropriate statistical test uses the binomial distribution. With samples as large as this one, however, the binomial distribution is well approximated by the normal distribution. For the continuous health measures, the "t" distribution is needed. But again, for large samples, the "t" distribution is well approximated by the normal distribution.

TABLE 1. Mean Values of Health Measures* in Cycle II Sample by Family Income and Race, and Infant Mortality and Birth Weight by Family Income and Race

Cycle II Health Measure	Cycle II Working Sample (n = 4777)	Blacks (n = 581)	Whites (n = 4196)	"z" Statistic for Black-White Difference	Adjusted "z" Statistic for Black-White Difference	Income < \$5,000/yr (n = 1645)	Income ≥ \$5,000/yr (n = 3132)	"z" Statistic for Low-High Income Difference	Adjusted "z" Statistic for Low-High Income Difference
ASSESS	0.4706	0.6127	0.4509	7.36	4.09	0.5939	0.4058	12.58	6.99
VISION	0.1160	0.1136	0.1163	0.20	0.11	0.0954	0.1268	3.21	1.78
BLOODP	0.0557	0.0671	0.0541	1.28	0.71	0.0614	0.0527	1.25	0.87
ALLERGY	0.1478	0.0878	0.1561	4.35	2.42	0.1015	0.1721	6.56	3.65
TENSION	0.4593	0.3373	0.4762	6.32	3.51	0.4304	0.4745	2.90	1.61
ABNORMALITY	0.0373	0.0430	0.0365	0.78	0.43	0.0389	0.0364	0.44	0.24
PERI†	−0.0392	−0.0736	0.0344	1.19	0.66	0.0711	−0.0971	7.48	4.16
ABSENT‡	0.0445	0.0422	0.0449	0.28	0.16	0.0483	0.0426	0.87	0.48
Other Health Measures	Total Population 1963-65	Black Population 1963-65	White Population 1963-65	Population With Income < \$5,000/yr. 1963–65	Populaltion With Income ≥ \$5,000/yr. 1963-65				
Infant mortality§	23.0	39.5	20.8	28.4	18.1				
Per cent of births with weight < 2,500 grams	7.87	14.01	7.01	8.89	7.05				

* Health measures defined in text.

† The mean of this variable is not zero, because standardization was done using the means and standard deviations for the entire Cycle II sample rather than for our working subsample.

‡ Means derived from subsample of working sample for which a school report was available (n = 4,333 rather than 4,777)

§ Deaths per 1,000 live legitimate births.[26]

this analysis such information was not provided by the Health Examination Survey (HES), in order to preserve the confidentiality of the Cycle II data. We obtain an approximation of the true standard errors for the Cycle II sample by adapting the results of a Monte Carlo study of data from the Cycle I HES survey (which utilizes the same sample design as Cycle II) reported by McCarthy.[27] McCarthy finds that the true standard error of sample means in the Cycle I data is, on the average, 1.8 times the value computed under the assumption that the data came from a simple random sample. He also estimates that the true standard error of regression coefficients computed from the Cycle I data is, on the average, 1.35 times the standard error computed under the assumption of simple random sampling. These two adjustment factors are applied to the estimates in an effort to control for the effects of the Cycle II sample design.

Statistically significant race differences (at the 5 per cent level, which is used throughout this paper) are reported for only three of the eight health measures: ASSESS, ALLERGY and TENSION. Black parents assess their children to be in poorer overall current health than do white parents. At the same time, black parents are less likely than white parents to report that their children are extremely tense or are afflicted with allergies. These three health measures are all subjective, since they are based on the parents' report. For the more objective measures, no race differences are statistically significant. The reported differences do suggest, however, that the black children in the sample are more likely to have physical abnormalities and high blood pressure than the white children. The data also suggest that black children are less likely to have periodontal disease and poor vision. In addition, black children are less likely than white children to be absent from school because of illness.

TABLE 2. Percentage of Black Children, Percentage of Children from Low-Income (< $5,000/yr) Families, and Mean Values of Health Measures in Cycle II Total and Working Samples*

Cycle II Measure	Total Sample	Working Sample
Percentage of black children	13.9	12.2
Percentage of children from low-income families	37.1	34.4
Health Measures		
ASSESS	0.4790	0.4706
VISION	0.0950	0.1160
BLOODP	0.0510	0.0557
ALLERGY	0.1580	0.1478
TENSION	0.4560	0.4593
ABNORMALITY	0.0420	0.0373
PERI	0.0000	−0.0392
ABSENT	0.0440	0.0445

* In computing the percentage or mean of a given measure in the total sample, observations with missing data for that measure alone are deleted.

To summarize, race comparisons of children's health based on these eight measures clearly yield a much less uniform impression than is drawn from inspection of data on infant mortality and low birth weight. Rather than exhibiting the dramatically large health deficits of black infants, older black children do not appear to have a consistently different health profile than their white counterparts. Moreover, those differences that are statistically significant only pertain to subjective measures. This divergence between the relative health profile of blacks in infancy and in mid-childhood is not simply a result of differences in data sources or differences between sample data and population data. Data for the Cycle II children regarding their infancy (birth weight, the incidence of congenital abnormalities and the parents' retrospective assessment of the infant's health) indicate that even the black children in the Cycle II sample had significantly poorer health in infancy than did the white children in the sample.

Income differences in these health measures are of the same order of magnitude as are the race differences, only three being statistically significant.**[28] Two of these place children from higher income families in better health (the parents' overall assessment and the periodontal index), while the allergies measure places them in worse health. These income-related differences are independent of race and are evident in the white sample as well (Table 3). It is notable that for the doctor's overall assessment of significant abnormalities there is no significant income difference.

Data for the period 1971-72 from Cycle I of the Health and Nutrition Examination Survey, however, give a somewhat different result regarding income differences in visual acuity when children are examined with their "usual correction"[29] In particular, a smaller proportion of low-income children had usual distance vision in the better eye of 20/20 or better, compared with high-income children; and a larger proportion of low-income children had usual distance vision of 20/50 or worse. It should be noted, however, that these relationships are not uniform over the three income classes studied. A larger proportion of low-income (less than $5,000) children had vision of 20/20 or better, compared with middle-income ($5,000–9,999) children. A larger proportion of high-income ($10,000 and over) children had vision of 20/50 or worse, compared with low-income children. Moreover, tabulations for the classification 20/30 or worse are not shown.

What conclusions can be drawn from this simple comparison of means between the black–white and low-income/high-income subsamples? First, neither race nor income differences are pronounced. Second, those variables for which there are significant race and income differences are, for the most part, subjective measures. Finally, and most notably, we report that for the overall evaluation of the child's health by the Public Health Service physician (ABNORMALITY), there are no significant differences either by race or by income class.

Our results with respect to race and income differences in the parent's assessment of the child's overall health are similar to those reported by the NCHS,[23] in spite of our use of a smaller sample. Our results, with respect to significant acquired abnormalities, differ from those of the NCHS. Both analyses indicate fewer abnormalities in white children, compared with black children, and in children from high-income families, compared with children from low-income families. But the NCHS differentials are statistically significant, while those in Table 1 are not, probably because we exclude congenital abnormalities, while the NCHS includes them. The black–white differential in the proportion of congenital abnormalities is significant in our sample.

Decomposition of Observed Race and Income Differences

We have documented that race and income differences in health status in the Cycle II sample children are much less sharp that the corresponding differences in measures of infant health. Nevertheless, some differences still do exist. To what ex-

---

[**] Note that the differences in the prevalence rates of tension and abnormal uncorrected distance vision between the high-income and low-income samples are significant if the unadjusted "z" statistics in Table 1 are used. Kish[28] shows that the procedure used to adjust for design effects overstates the true standard error of the difference in means. This is because "... two subclass means from clustered samples, although based on distinct sets of elements, tend to come from the same set of clusters. The positive correlation between cluster influences on the two means tends to reduce the variance of the difference (p. 582)." Kish also shows that the estimate of the standard error of the difference in means based on the assumption of simple random sampling understates the true standard error. Hence, caution is required with respect to conclusions concerning the statistical significance of gross income difference in the prevalence of tension and abnormal vision. Note that the preceding point refers only to computing the standard error of the difference in means and is not relevant to our procedure for computing adjusted "z" statistics of regression coefficients.

TABLE 3. Gross and Net Income Differences in the Health Status of White Children*

Health Measure	High Income Mean	Low Income Mean	Gross Income Difference†	"Z" Statistic on Gross Difference	Net Income Difference‡	Ratio of Net to Gross Income Difference
ASSESS	0.394	0.589	−0.195§	11.62	−0.079	0.41
VISION	0.127	0.091	0.036	3.25	0.006	0.17
BLOODP	0.052	0.056	−0.004	0.84	0.006	−1.50
ALLERGY	0.176	0.108	0.068§	5.59	0.016	0.24
TENSION	0.484	0.456	0.028	1.65	−0.002	−0.07
ABNORMALITY	0.036	0.037	−0.001	0.04	−0.001	1.00
PERI	−0.097	0.117	−0.214§	8.38	−0.077	0.36
ABSENT	0.043	0.049	−0.006	0.85	0.001	−0.17

* There are 2,718 white children in the high-income (≥ $5,000/yr) sample and 1,227 white children in the low-income (< $5,000/yr) sample.

† This is computed as the mean in the high-income sample minus the mean in the low-income sample.

‡ Computation described in text.

§ Significant income differences at the 5 per cent level of significance, whether or not adjustment is made for sample design.

tent are these uniquely associated with income and race, and to what extent can they be attributed to correlated socioeconomic factors? This question is especially relevant here because it is the subjective measures for which significant differences are observed, and these measures are most likely to be affected by systematic variations in reporting and/or scaling related to socioeconomic characteristics.

Race Differences

It is well known that race and income are highly correlated. Thus it is not surprising that two of three of the health measures displaying significant race differences also exhibit significant income differences (Table 1). An obvious first step, therefore, is to try to determine whether these observed race differences really are just a result of differences in income. To do so we obtain mean values for the eight health measures when the Cycle II sample is cross-classified by both income and race (Table 4). There are still significant race differences in health, but different measures are affected in the two income classes. Among low-income families significant race differences are observed for tension and the periodontal index, and for both of these measures black children are rated in better health than white children from families of comparable low-income levels. In high income families, significant race differences are reported for the parents' assessment, allergies and tension, black children being rated in worse health according to the parents' assessment and in better health according to tension and allergies. Thus, within income classes significant health differences still exist between black and white children. As before, these differences do not uniformly favor children of either race, but most of the significant differences are for the subjective measures and most of them show blacks to have a better rating.

These results can again be contrasted with comparable data for the two infant health measures discussed earlier. Even within income classes, race differences in infant mortality and in the incidence of low birth weight remain large and consistently favor whites (Table 5).

TABLE 4. Gross and Net Race Differences in Children's Health Status

Health Measure	Gross Difference (Black Mean-White Mean)	Low-Income Families (< $5,000/yr) Black Mean (n = 418)	White Mean (n = 1,227)	Difference	"z" Statistic for Black-White Difference	High-Income Families (> $5,000/yr) Black Mean (n = 163)	White Mean (n = 2,969)	Difference	"z" Statistic for Black-White Difference	Net Difference
ASSESS	0.162*	0.612	0.589	0.023	0.89	0.614	0.394	0.220*	5.57	0.119†
VISION	−0.002	0.108	0.091	0.017	0.98	0.129	0.127	0.002	0.10	0.009‡
BLOODP	0.013	0.069	0.056	0.013	0.79	0.061	0.052	0.009	0.51	−0.003‡
ALLERGY	−0.068*	0.084	0.108	−0.024	1.39	0.098	0.176	−0.078*	2.57	0.205†
TENSION	−0.139*	0.354	0.456	−0.102*	3.66	0.295	0.484	−0.189*	4.74	−0.153*‡
ABNORMALITY	0.007	0.046	0.037	0.009	0.80	0.037	0.036	0.001	0.03	0.018‡
PERI	−0.039	−0.063	0.117	−0.180*	3.63	−0.101	−0.097	−0.004	0.10	−0.059‡
ABSENT	−0.003	0.045	0.949	−0.004	0.32	0.034	0.043	−0.009	0.51	−0.005‡

* Significant race differences in means at the 5 per cent level of significance. These results hold whether or not an adjustment for the sample design is made in the variances with the exception of ALLERGY in the high-income families, in which case the adjusted "z" is 1.43.

† Net difference equals difference in health levels predicted for blacks and whites assuming that all explanatory variables are fixed at their mean values for the full working sample. We do not present the results of significance tests in these cases because of computational difficulty.

‡ Net difference equals regression coefficient of a race dummy variable (1 = black) from a pooled regression of black and white children that holds constant all other independent variables.

A further way of investigating the nature of race differences in children's health is to look at residual race differences after a much larger list of socioeconomic variables is held constant. This is done with the use of multiple regression analysis.†† The dependent variables in the regression equations are the eight health status measures. For the explanatory variables a set suggested by the economic model of family investment in children's health described in Edwards and Grossman was used.[30] Included are family income; parents' educational attainment; whether the child's father lives with the family; whether the child is a twin or a first-born; whether a foreign language is spoken in the home, an indication of the region of residence and size of city of residence; and the sex of the child (the latter is included only for health measures that are not standardized by sex). These variables are defined in detail in Edwards and Grossman.[30]

Race differences in children's health net of differences in this set of socioeconomic variables are presented in the last column of Table 4. Net differences are computed in two different ways, depending on whether or not there are significant race differences in slope coefficients in the underlying health equations. If there are no significant differences in the slope coefficients by race (as in the case for the dependent variables, BLOODP, VISION, TENSION, ABNORMALITY and ABSENT), the net race difference is represented by the regression coefficient of a race dummy variable (black = 1) from a pooled regression of black and white children that holds constant all other dependent variables. When there are significant race differences in the slope coefficients (as in the case of the dependent variables PERI, ASSESS and ALLERGY), the net difference is computed according to the following formula: Net difference = $(a_b + b_b \overline{X}) - (a_w + b_w \overline{X})$, where a_b and b_b represent the intercept and vector of slope coefficients estimated when one is using the black subsample, a_w and b_w represent the corresponding coefficients estimated when one is using the white subsample, and $\overline{X}$ represents a vector of means of the set of explanatory socioeconomic variables computed over the entire working sample. In either case, the net difference can be interpreted as that portion of the gross difference that would remain if both race groups had the same mean values of all explanatory variables.

In general, the net differences are not smaller than the gross differences. Indeed, in some cases they are substantially larger and /or change signs. For example, for the tension variable, the net difference is greater than the gross difference, implying that if both black and white families had the same socioeconomic characteristics, the blacks relatively would be even less likely than whites to report excessive tension. Or, in the case of allergies, the net difference implies that blacks would be more likely than whites to report allergies if both blacks and whites had the same

TABLE 5. Infant Mortality and Birth Weight

	Low Income (< \$5,000/yr.)		High Income (≥ \$5,000/yr.)	
	Blacks	Whites	Blacks	Whites
Infant mortality	43.9	24.2	23.6	18.6
Incidence of low birth weight	13.2	7.6	17.1	6.6

Sources and definitions of health measures are the same as in Table 1.

---

†† Ordinary least squares was used to estimate linear equations using the health measures as dependent variables. When the health measures are dichotomous, this procedure is not correct because the residuals variance will exhibit heteroscedasticity. However, experiments with a more appropriate (and much more costly) estimation procedure—LOGIT—indicated that our results are virtually unchanged when LOGIT is used in place of ordinary least squares.

socioeconomic characteristics. The net difference for the ABNORMALITY variable is also greater than the gross difference, but it still does not increase enough to reach statistical significance. The one important exception to the generalization above is the parents' assessment variable. In this case, the net difference is less than the gross difference.

To sum up, the various statistics in Table 4 suggest the following conclusions. First, our view of race differences in children's health is not altered substantially when income and other socioeconomic variables are held constant. Second, these differences do not obviously favor children of either race. Third, it is notable that for some of the harder measures—the physician's overall assessment, unusually high blood pressure, poor vision—significant race differences are never reported. On the other hand, differences in subjective measures tend to be significant. The latter may reflect actual differences or it may reflect only perceived differences. It may even reflect systematic differences in scaling: for example, a family undergoing many stresses (as a low-income family might) may not place much weight on a child's symptoms of allergy or tension. Indeed, it may only be a visit to a doctor that makes a family aware that a child is afflicted with allergies or is excessively tense. Thus, it is clear that results for these subjective measures must be interpreted with extreme caution.

Income Differences

Gross race differences in children's health did not disappear when various socioeconomic factors (including income) were held constant. Does the same conclusion hold for income-related differences? In other words, to what extent do the gross differences in health status between income classes disappear when additional socioeconomic factors are held constant? To answer this question, we use the same type of multiple regression analysis described earlier. We simplify the analysis, however, by restricting the decomposition of gross income differences to the white sample only.

According to the adjusted "z" statistic, the list of health measures for which there are significant income differences is the same for the white sample as it is for the full black-and-white sample. If the unadjusted "z" statistic is used, however, there is no significant income difference for tension in the white sample, but there is in the white–black sample.

Gross and net income differences for the eight health measures are presented in Table 3. The net income differences are computed similarly to the corresponding net race differences. They should be interpreted as the difference in mean health status between the two income classes if all of the socioeconomic variables (other than income, of course) took the same values in both classes. Unlike the case of race differences, there were no significant differences for health measures in the slope coefficients in the two income classes. The calculations, however, do permit the coefficient of income to be different among high- and low-income families. The theoretical rationale for this is discussed in Edwards and Grossman.[30] The net income difference for any health measure is computed as the coefficient of income multiplied by the difference in mean income in the two income classes. Note that in these equations income is measured as a continuous variable, whereas race is dichotomous. Income in Cycle II is reported in ten classes: less than $500, $500–$999, $1,000–$1,999, $2,000–$2,999 $3,000–$3,999, $4,000–$4,999, $5,000–$5,999, $6,000–$6,999, $7,000–$9,999, $10,000–$14,999, and $15,000 and over. We convert these classes into a continuous variable by assigning midpoints to the closed intervals, $250 to the lowest interval and $20,000 to the highest interval.

The number of health status measures for which statistically significant income differences exist is reduced from three (ASSESS, ALLERGY and PERI) to zero when related socioeconomic factors are held constant. For these three (and for most of the other measures as well) the magnitude of the net difference is substantially smaller than the corresponding gross difference (see the last two columns of Table 3). For example, on the basis of the gross difference, about 20 per cent more of the high-income parents than of the low-income parents assessed their children's health as being very good. When related socioeconomic factors are held constant, this differential is reduced to only 8 per cent. To conclude, no significant income differences in health status are reported when related socioeconomic characteristics are held constant, and the magnitudes of all of these income differences are greatly diminished. The main conclusion to be drawn is clear: gross income differences in health greatly overstate the true relationship between family income and health.

Further insight into the nature of gross income differentials in health status is obtained by study of the precise role of the explanatory socioeconomic variables. Table 6 presents calculations which illustrate how the gross income differences are decomposed among the various explanatory factors for the three health variables that exhibited significant gross income differences in Table 3. The procedure simply is to multiply the coefficients of these explanatory variables by the differences in their mean values in the high- and low-income samples of children. Several results in Table 6 are noteworthy. First, more than 80 per cent of the differences in the four health measures between the high- and low-income subsamples can be accounted for by differences in the independent variables that we have included in our multiple regression equation. Second, a detailed examination of the decomposition indicates that differences in parents' average schooling between high- and low-income families account for about as much or even more of the gross differences as does income itself. And finally, more of the income-related gross differences in children's health are accounted for by differences in correlated socioeconomic and regional characteristics of the child than by family income itself.

Our findings with respect to race and income differences may be compared with those in a recent study of infant mortality

TABLE 6. Components of the Difference in ASSESS, ALLERGY and PERI Between White Children from High- and Low-Income Families*

Component	ASSESS	ALLERGY	PERI
Family income	−0.079	0.016	−0.077
Parents schooling	−0.071	0.045	−0.131
Other family characteristics†	0.001	0.004	−0.011
Characteristics of child‡	−0.003	0.003	0.001
Region	−0.025	−0.005	0.019
City size	−0.005	−0.004	−0.001
Total = predicted gross difference	−0.182	0.059	−0.200
Actual gross difference	−0.195	0.068	−0.214

* Computed from coefficients in a multiple regression which includes the explanatory variables listed in the text.

† These include whether or not a foreign language is spoken in the home and whether or not the father is absent from the home.

‡ These include whether or not the child is a first-born or a twin, and his or her sex. The latter is not included for PERI which is standardized by sex.

by Gortmaker.[31] He attempts to determine what portion of the large income and race differences in infant mortality can be explained by differences in parents' educational attainment, mother's age, the child's birth order and the previous pregnancy experience of the mother. He finds that poor families and black families still display a much higher incidence of infant mortality even when these factors are controlled for. This contrasts with the findings in this article for later childhood. Once again, we are struck by the different views of child health from the alternative perspectives of infancy and midchildhood.

Conclusions

The first point that must be made in concluding is that income and race differences in infant mortality provide a poor and even misleading description of income and race differences in the health of older children for the 1963-65 period. This suggests that while it may be appropriate for one to use infant mortality statistics in broad across-country comparisons of the health status of various populations, one should not use these statistics to indicate the relative health statuses of various groups of older children within the United States.

We do find differences in the health status of black and white children and of children from high- and low-income families, but these differences by no means overwhelmingly favor the white or high-income children. With respect to differences by race, whether or not they are adjusted for differences in associated socioeconomic factors, significant differences exist primarily for the subjective health measures, and these do not always show black children to be rated in poorer health than their white counterparts. In the case of income differences in health, the high-income children do appear to be rated healthier according to most measures, but their advantage is diminished greatly when one controls for related socioeconomic factors like parents' educational attainment. It is important that one notes, however, that with respect to the variable that might reflect the most serious health problems—whether or not the doctor reports the child to have a "significant abnormality"—there are never statistically significant differences either by race or by income.

These results are especially intriguing when one considers that the children in the Cycle II sample were growing up in the late 1950s and early 1960s, a period prior to the introduction of a number of government health programs targeted at the poor such as Medicaid, maternal and infant care projects, and community health centers. A natural follow-up to our study would be similar analyses using a more recent national sample of children.

Implicit in our discussion is the necessity for recognizing the multidimensional nature of health. Our results clearly differ for different health status measures. Poor and black children tend to be rated in better health when subjective health measures are studied, but they are rated as being in the same or worse health when more objective health measures are used. Thus, our findings underscore the importance of treating children's health status as multidimensional, and illustrates how the use of a single health index could lead to erroneous conclusions about health status and its relation to income and race. Of course, when a set of measures is used instead of a single measure, the researcher is faced with more difficulty in synthesizing his results. But the fact that one is unlikely to be able to draw neat conclusions from a multifaceted analysis of health is not an acceptable reason to restrict that analysis to a single health measure.

The primary purpose of this article has been to expand upon commonly held notions about income and race differences in children's health by the use of a study of children who were 6 to 11 years old in 1963-65. We have found that the unambiguous picture drawn from infant mortality statistics becomes fuzzy and difficult to decipher when a collection of morbidity measures provides the medium. This raises an important and, as yet, unanswered question, especially if our findings are reinforced by other studies using different measures and different time periods. Why are the income and race differences in infant mortality so striking when corresponding differences in mid-childhood morbidity are not? This divergence is not simply a result of the type of measure (mortality vs. morbidity); differences in the mortality of older children also are less striking. Perhaps the health differences between various demographic groups are minimized in the mid-childhood years. If so, we need to know why. The answer to these questions are clearly pertinent to the conduct of public policy toward the welfare of children.

Acknowledgments

The following physicians gave us helpful advice: John McNamara, M.D., then Assistant Professor of Public Health and Pediatrics at Columbia University School of Public Health and Associate Commissioner in the New York City Department of Health; Roy Brown, M.D., Associate Professor of Community Medicine and Pediatrics at the Mount Sinai School of Medicine of the City University of New York; Thomas Travers, D.D.S., Director of Ambulatory Care in the New York City Department of Health; and Ruth T. Gross, M.D., Professor of Pediatrics and Director of Ambulatory Pediatrics, Stanford University Medical Center, Stanford, California. We are also indebted to Mary Grace Kovar for her helpful comments on an earlier draft.

References

1. Newberger E, Newberger CM, Richmond JB. Child health in America: toward a rational public policy. Milbank Mem Fund Q 1976;54:249

2. Keniston K. All our children: The American family under pressure. New York: Harcourt Brace Jovanovich, 1977.

3. National Center for Health Statistics. Plan, operation, and response results of a program of children's examination. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 1, No. 5, 1967.

4. U.S. Bureau of the Census. 1970 census of population, Vol I, characteristics of the population, Part 1, U.S. summary, sections 1 and 2. Washington, D.C.: U.S. Government Printing Office, 1973.

5. Sullivan DF. Conceptual problems in developing an index of health. U.S. Department of Health, Education and Welfare, Public Health Service, Series 2, No. 17, 1966.

6. Berg RL. Health status indexes. Chicago: Hospital Research and Educational Trust, 1973.

7. Grossman M. The demand for health: a theoretical and empirical investigation. New York: Columbia University for the National Bureau of Economic Research, 1972.

8. Grossman M. The correlation between health and schooling. In: Terleckyj NE, ed. Household production and consumption. New York: Columbia University Press for the National Bureau of Economic Research, 1975:147.

9. Birch HG, Gussow JD. Disadvantaged children: health, nutrition, and school failure. New York: Harcourt, Brace and World, 1970.

10. Starfield B. Stability and change: another view. Am J Public Health 1981;71:301.

11. Wallace HM. Health services for mothers and children. Philadelphia: WB Saunders, 1962.

12. Mechanic D. The influence of mothers on their children's health, attitudes and behavior. Pediatrics 1964;33:444.

13. Talbot NB, Kagan J, Eisenberg L, eds. Behavioral science in pediatric medicine. Philadelphia: WB Saunders, 1971.

14. Kaplan RS, Lave LB, Leinhardt S. The efficacy of a comprehensive health care project: an empirical analysis. Am J Public Health 1972;62:924.

15. Hu T. Effectiveness of child health and welfare programs: a simultaneous equations approach. Socio-Economic Planning Sciences 1973;7:705.

16. Kessner DM. Assessment of medical care for children, vol 3, Contrasts in health status. Washington, D.C.: Institute of Medicine, 1974.

17. Haggerty RJ, Roghmann KJ, Pless IB. Child health and the community. New York: John Wiley and Sons, 1975.

18. Inman RP. The family provision of children's health: an economic analysis. In: Rosett R, ed. The role of health insurance in the health services sector. New York: Neale Watson Academic Publications, 1976:215.

19. Starfield B. Health needs of children. In: Children's medical care needs and treatment. Harvard Child Health Project Report Vol II. Cambridge: Ballinger Publishing, 1977:23.

20. National Center for Health Statistics. Binocular visual acuity of children: demographic and socioeconomic characteristics. U.S. Department of Health, Education and Welfare, Public Health Ser-

vice, Vital and Health Statistics, Series 11, No. 112, 1972.

21. Kelly JE, Sanchez M. Periodontal disease and oral hygiene among children. National Center for Health Statistics, U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 11, No. 117, 1972.

22. Russell AL. A system of classification and scoring for prevalence surveys of periodontal disease. Journal of Dental Research 1956;35:350.

23. National Center for Health Statistics. Examination and health history findings among children and youths. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 11, No. 129, 1973.

24. Koran LM. The reliability of clinical methods, data and judgments, part one. N Engl J Med 1975;293:642.

25. Koran LM. The reliability of clinical methods, data and judgments, part two. N Engl J Med 1975;293:695.

26. National Center for Health Statistics. Infant mortality rates: socioeconomic factors. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 22, No. 14, 1972.

27. McCarthy PJ. Pseudoreplication: further evaluation and applications of the balanced half-sample technique. National Center for Health Statistics, U.S. Department of Health, Education, and Welfare, Public Health Service, Vital and Health Statistics, Series, No. 31, 1969.

28. Kish L. Survey sampling. New York: John Wiley and Sons, 1965.

29. National Center for Health Statistics. Monocular visual acuity of persons 4–74 years. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 11, No. 201, 1977.

30. Edwards LN, Grossman M. Children's health and the family. In: Scheffler RM, ed. Advances in health economics and health services research. Greenwich, Conn.: JAI Press, 1981:35.

31. Gortmaker SL. Poverty and infant mortality in the United States. Am Sociol Rev 1979;44:280.

Erratum

The figures on Table 1, page 181, of "Predicting the Outcome of Primary Care," by D. D. Wright and R. L. Kane, *Medical Care*, February 1982, are incorrect. The predictive value of a positive test should be 0.785, and the predictive value of a negative test should be 0.571.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Regional Differences in Hospital Use Standardized by Reported Morbidity

MARK S. BLUMBERG, M.D.

The purpose of the study was to determine whether regional differences in hospital use parallel regional morbidity differences. It showed that they do not. We used the 1975 U.S. Health Interview Survey Metropolitan sample under age 65 to analyze the nonobstetric hospital use and the reported morbidity in the four U.S. regions. The West sample used less hospital care than those in other regions, even though other regions reported less morbidity. We indirectly standardized hospital use by age-sex specific morbidity. The standardized ratio of actual to expected hospital nights in the West was half that of the North Central Region. There were smaller regional differences in surgical hospital use than for nonsurgical hospital use. In 1975 the total U.S. metropolitan population under age 65 would have used one third fewer nonobstetric hospital nights if it had followed Western practices.

THE REGIONS of the United States differ in age-sex specific hospital use per person. In 1975 the annual number of hospital days per thousand persons was lowest in the west for each of the eight age-sex categories in Table 1. The rates in the west ranged from 57 to 67 per cent of the corresponding north central rates, which were the highest for those under age 65.

Almost all hospital care is for treatment or diagnosis of health problems. Hence the health status of a population is a key determinant of its relative need for hospital care. While age and sex commonly are used as proxies for health status, we sought more direct measures of population health by region to determine how many of the hospital use differences were attributable to health differences.

Data Source

Each year since 1958, the Bureau of the Census has conducted the U.S. Health Interview Survey (HIS) for the Public Health Service. The survey is designed and analyzed by the National Center for Health Statistics. We obtained public use computer tapes of the 1975 HIS because we were interested in the private health care coverage of respondents and that year information on this was obtained.

The total annual HIS sample of about 112,000 persons in 41,000 households is representative of the U.S. civilian-noninstitutional population. The sample interviews are staggered throughout the calendar year, thus avoiding seasonal bias. However, the ample process results in substantial understatement of both morbidity and hospital use for the total population age 65 and over. Those in older age groups are more likely than those under age 65 to be institutionalized for health

---

From the Kaiser Foundation Health Plan Inc., Oakland, California.

Reprint requests: Mark S. Blumberg, M.D., Kaiser Foundation Health Plan, Inc., One Kaiser Plaza, Oakland, CA 94612.

0025-7079/82/0900/0931/$01.50 © J. B. Lippincott Co.

problems. Older persons also have a higher mortality rate. Since the health and hospital experience of the institutional population and of decedants is not included in the Health Interview Survey, we chose to analyze the HIS sample population that was under age 65. The study was limited further to those residing in the over 200 Standard Metropolitan Statistical Areas (SMSAs) that were part of the Health Interview Survey sample.

Health Status and Hospital Use Measures

The two basic measures of morbidity from the HIS Survey are the "restricted activity day" and conditions causing a "limitation of activity." A day of restricted activity is one on which a person cuts down on his usual activities for the whole of that day because of an illness or an injury." Data on restricted activity days are collected for the two-week period prior to the interview. They include bed disability days, which in turn include hospital days. By definition, limitations of activity are caused by chronic conditions defined as those on a stipulated list or having been present for three months or more.

The data in Table 2 show that the 1975 West HIS Sample (including those not residing in SMSAs) was highest of any region in the number of restricted activity days per person per year in each age-sex category under age 65. The North Central region was frequently the region with the lowest number of restricted activity days per person. The pattern is similar for the percentage of the population with limitation of activity. These data suggest that the interregional differences in hospital use shown in Table 1 are not due to corresponding differences in morbidity by region.

The hospital use reported by the Health Interview Survey consisted of all that occurred in short-stay hospitals of any ownership or location (e.g., including federal and foreign) during the year prior to interview. While the study was limited to episodes in short-stay hospitals, it could have included some episodes in long-stay units of short-stay hospitals. To minimize the effect of such long stays we truncated the length of all sample episodes at 45 days.

TABLE 1. Non-Federal Short Stay Hospital Days* per 1,000 Population by Sex, Age and Region—U.S. National Hospital Discharge Survey[6] 1975

Sex and Age	Region				
	Northeast	North Central	South	West	All U.S.
Male					
Under 15	357	461·†	344	264°‡	365
15–44	650	789·	571	488°	634
45–64	1898	2035·	1492	1270°	1700
65+	4714	5065·	4075	3328°	4379
All	1215	1321·	1000	812°	1104
Female					
Under 15	281	366·	269	205°	287
15–44	805	1072·	830	658°	858
45–64	1840	2272·	1545	1528°	1792
65+	4385	4757·	3612	2988°	4016
All	1363	1570·	1158	946°	1280

* Excludes well newborns and obstetric deliveries.

† · = Highest value on each line.

‡ ° = Lowest value on each line.

TABLE 2. Restricted Activity Days per Person per Year by Region, Sex and Age—1975 Health Interview Survey[*][†]

Sex and Age	Region				
	Northeast	North Central	South	West	All
Male					
Under 6	13.1	11.8	11.2°[‡]	15.1˙[§]	12.5
6–16	10.5	9.5°	9.8	12.0˙	10.2
17–24	9.9	7.7°	10.0	14.5˙	10.2
25–44	12.1°	12.2	13.7	13.8˙	13.0
45–64	19.2°	19.2°	24.7	25.4˙	22.0
65+	29.0°	29.7	41.4˙	36.0	34.3
All	14.7	13.8°	16.6	17.8˙	15.6
Female					
Under 6	11.7	9.6°	10.4	16.4˙	11.6
6–16	11.6	10.5	9.0°	13.8˙	10.9
17–24	14.4	13.4°	13.7	16.8˙	14.3
25–44	16.1°	16.1°	19.5	20.7˙	18.1
45–64	23.6	21.3°	28.0	34.3˙	26.2
65+	33.6°	38.7	48.2˙	45.3	41.4
All	18.5	17.9°	20.8	23.6˙	20.0

[*] Data are based on household interviews of the civilian, noninstitutional population.

[†] Gentile A. Disability Day U.S. 1975: National Center for Health Statistics, Data from the National Health Survey: Series 10, No. 118 (DHEW Publication No. (PHS) 78-1546).

[‡] ° = Lowest value on each line.

[§] ˙ = Highest value on each line.

Obstetric deliveries are the most common cause of hospital admissions in the United States. However, hospitalization for obstetric deliveries is almost universal and hence, inclusion of obstetric deliveries could confound other hospital use rates that can be more discretionary. Furthermore, normal pregnancy is not considered morbidity by the HIS. For these reasons we excluded all hospitalization for obstetric deliveries, as well as for abortions. (By convention, the HIS source data omit all hospital care of well newborns.)

Standards for Hospital Use

We developed standards for hospital use for the 1975 HIS SMSA sample under age 65 (72,360 persons). These standards are based on hospital use, restricted activity days and limitation of activity due to chronic conditions of the U.S. SMSA sample under age 65. First the sample population was divided into those who had one or more conditions causing limitation of activity and those who did not. These two subsamples were then divided into eight age categories for males and for females, for a total of 32 cells. Finally, separate standards were derived for hospital episodes and for hospital nights.

Table 3 presents the standards used in this analysis. The value of 1.734 on the bottom line in the left column was derived as follows: there were 2,705 restricted activity days (in a 2-week period) reported by 1,002 males, age 55 to 64, who were limited in activity. This same sample reported 4,690 hospital nights in the past year. The standard is simply 4,690 nights divided by 2,705 restricted activity days, or 1.734 annual hospital nights per restricted activity day in 2 weeks.

The value of the standard for hospital episodes in this same population was based on the fact that this sample had 360 hospital episodes per 2,705 restricted activity days or 0.1331, as shown on the bottom of the second column.

TABLE 3. Annual Number of Hospital Episodes and Nights* per Restricted Activity Day (in a 2-week period:) by Age and Sex for Persons With and Without Limitation of Activity Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

	Sex			
	Male		Female	
Age	Nights	Episodes	Nights	Episodes
Persons not limited in activity				
Under 1	2.650	0.3376	2.000	0.2638
1–5	0.633	0.1605	0.546	0.1264
6–16	0.584	0.1265	0.432	0.1060
17–24	0.812	0.1600	1.772	0.1633
25–34	0.910	0.1636	1.119	0.1855
35–44	1.039	0.1728	1.511	0.2526
45–54	1.765	0.2434	1.679	0.2378
55–64	2.311	0.2582	1.472	0.1699
Persons limited in activity				
Under 1	11.882	0.5882	11.267	0.8667
1–5	2.123	0.2538	1.139	0.2025
6–16	1.096	0.1570	0.588	0.0785
17–24	1.550	0.1403	0.848	0.1168
25–34	1.429	0.1311	1.349	0.1479
35–44	2.071	0.1602	1.291	0.1238
45–54	1.693	1.1448	1.232	0.1155
55–64	1.734	0.1331	1.180	0.1126

* Excludes obstetric deliveries, abortion and well newborns.

We used these standards to derive estimates of the expected number of hospital episodes and hospital nights for each of the four U.S. regions. For example, in the Northeast region there were 226 males, age 45 to 54, limited in activity, who reported 435 restricted activity days. Their 435 of restricted activity days were multiplied by the corresponding U.S. standards (1.693 nights and 0.1448 episodes). This resulted in 736.22 expected nights and 62.97 expected episodes in their cell. The expected nights and episodes obtained in a like way for all of the 64 cells were summed for each of the four regions. These expected nights and episodes were then compared with the actual number of hospital nights and episodes in each region.

Results of Analysis

Some of the morbidity characteristics of the study samples, by region of the country, are shown at the top of Table 4. Line d shows that the percentage of the sample limited in activity varied by region with the West being the highest and the North Central region being the lowest (12.05 per cent versus 9.18 per cent). Lines e and f show that those who were limited in activity had many more restricted activity days per person per year than those who were not limited in activity. While this was true for each region, those in each category in the West had the highest number of restricted activity days per person, while those in the north cental region had the fewest number of restricted activity days per person.

Because of the greater reported morbidity in the West, this region also had the highest annual expected hospital episodes (line j) and hospital nights (line 1). Conversely, the North Central region had the lowest annual number of expected hospital episodes and nights per hundred persons. Table 5 presents the indirectly stand-

TABLE 4. Selected Morbidity and Expected Hospital* Use Measures for Persons With and Without Limitation of Activity by Region Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

Selected Measures	Region				
	Northeast	North Central	South	West	All
Sample Persons					
a Not limited in activity	17,482	17,296	17,246	12,884	64,908
b limited in activity	2,024	1,749	1,912	1,767	7,452
c All	19,506	19 045	19,158	14,651	72,360
d Per cent of samp'e with limitation of activity	10.38	9.18°†	9.98	12.6 ·‡	*10.30*
Annual restricted activity days per person					
e Not limited in activity	10.10	9.99°	10.31	12.73·	10.65
f Limited in activity	56.35	55.67°	62.47	66.82·	60.24
Annual expected hospital use per 100 persons					
Episodes					
g Not limited in activity	6.80	6.74°	7.07	8.44·	7.18
h Limited in activity	28.37	27.59°	31.55	33.34·	30.18
i All	9.04	8.65°	52	11.44·	9.55
Nights					
j Not limited in activity	40.99	39.88°	42.33	50.04·	42.85
k Limited in activity	306.99	299.76°	341.52	354.54·	325.43
l All	68.59	63.75°	72.19	86.76·	71.95

* Excludes obstetric deliveries, abortions and well newborns.

† ° = Lowest value on each line.

‡ · = Highest value on each line.

§ Expected base on age, sex and morbidity.

ardized ratios for hospital episodes, nights and lengths of stay by region of the country. Each ratio in Table 5 is the observed (i.e., actual) value divided by the corresponding expected value for a particular sample population. The ratio for the total U.S. is 1.0 by definition. A ratio of over 1.0 for a sample indicates that its actual use was greater than its expected use (based on U.S. standards). A value of under 1.0 indicated that the sample's actual use as less than expected. Line e gives the indirectly standardized ratio for all hospital episodes. It is highest in the North Central region (1.175) and lowest in the West (0.786). Given the same reported age, sex and morbidity experience in each region, this indicates that if the indirectly standardized hospital episode ratio of the West had been applied to the U.S. as a whole, total U.S. hospital admission rates would have been 21 per cent lower (i.e., 100[1–0.786]). The lowest episode ratio is for males in the West not limited in activity (line a, 0.748). However, the West has the lowest ratio for lines a, b, c and d which are the most disaggregated data presented. The least regional variation in episode ratios is for females who are limited in activity (line d).

The indirectly standardized ratios for hospital nights are more varied than the ratios for hospital episodes. Thus the overall West ratio for hospital nights (line j) is only 0.644. This indicates that if all hospital use in the U.S. had paralleled the West, total U.S. hospital nights would have been 35 per cent less than observed! The West has the lowest hospital night ratios for each

TABLE 5. Indirectly Standardized Ratios* for Hospital Episodes, Nights, And Length of Stay† for Persons With and Without Limitation of Activity by Sex and Region Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

Indirectly Standardized Ratio (Actual to Expected)	Region			
	Northeast	North Central	South	West
Episodes				
Not limited in activity				
a Male	0.911	1.200˙‡	1.115	0.748°§
b Female	1.000	1.167˙	1.059	0.763°
Limitation of activity				
c	1.229˙	1.174	0.927	0.802°
d Female	0.925	1.145˙	1.074	0.878°
All				
e Both sexes	0.980	1.175˙	1.057	0.786°
Nights				
Not limited in activity				
f Male	0.937	1.294˙	1.068	0.659°
g Female	1.126	1.186˙	1.053	0.619°
Limited in activity				
h Male	1.217	1.355˙	0.839	0.634°
i Female	1.085	1.304˙	1.007	0.670°
All				
j Both sexes	1.094	1.276˙	0.994	0.644°
Length of stay				
Not limited in activity				
k Male	1.028	1.078˙	0.958	0.881°
l Female	1.126˙	1.016	1.044	0.813°
Limited in activity				
m Male	0.990	1.154˙	0.905	0.791°
n Female	1.173˙	1.135	0.938	0.763°
All				
o Both sexes	1.164˙	1.086	0.940	0.819°

* Standardized for age, sex and morbidity.

† Excludes obstetric deliveries, abortions and well newborn.

‡ ˙ = Highest value on each line.

§ ° = Lowest value on each line.

of the disaggregated cells (lines f, g, h and i), while the North Central area has the highest ratios for these cells.

The data on line o are the indirectly standardized ratios for lengths of hospital stay by region of the country. The West has the lowest length-of-stay ratio (0.819), while the ratio in the Northeast is the highest (1.164). However, the North Central area also has a ratio above 1.

The extremely low hospital night ratio of the West (0.644) is due almost equally to a lower episode ratio (0.786) and to a lower length-of-stay ratio (0.819) for that region. In contrast, about two thirds of the high night ratio for the North Central region (1.276) is due to a high episode ratio (1.175) and about one third due to a high length-of-stay ratio (1.086).

The actual annual hospital use rates and lengths of stay for our HIS sample are given in Table 6, separately for cases which were and were not surgically treated. There was much more variation by region in these rates for cases that were not treated surgically (lines b, e and h) than for

TABLE 6. Actual Annual Hospital Episodes and Nights* per 100 Persons and Average Length of Stay Surgically vs. Non-Surgically Treated by Region Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

Hospital Use Measures	Region				
	Northeast	North Central	South	West	U.S.
Annual hospital episodes					
Per 100 persons activity					
a. Surgically treated	4.65°†	5.30	5.48˙‡	5.04	5.12
b. All other	4.21	4.86˙	4.58	3.95°	4.43
c. Total	8.86	10.17	10.06	9.00	9.55
Annual hospital nights					
Per 100 persons activity					
d. Surgically treated	35.30	37.91˙	37.15	30.24°	35.45
e. All other	39.75	43.40˙	34.63	25.60°	36.49
f. Total	75.05	81.31˙	71.78	55.85°	71.95
Average length of stay (nights)					
g. Surgically treated	7.59˙	7.15	6.78	6.00°	6.92
h. All other	9.43˙	8.93	7.57	6.48°	8.24
i. Total	8.47˙	8.00	7.13	6.21°	7.53

* Omits obstetric deliveries, abortions and well newborns.

† ° = Lowest value on each line.

‡ ˙ = Highest value on each line.

cases were treated surgically (lines a, d and g). For example, the annual hospital nights per 100 persons per year in the West for cases that were not treated surgically was only 60 per cent of the figure for the north central region) 25.6 vs. 43.4). The annual surgical episode rate in the West is almost equal to the U.S. rate and the rate is only a little higher in the North Central region. The average stay of nonsurgical cases in the Northeast is 45 per cent higher than it is in the West.

Validity of the Data and Method

Because the findings above were dramatic and unexpected, we were concerned about the validity and applicability of the morbidity data used. The marked interregional variations of hospital use in the Health Interview Survey are confirmed by the Hospital Discharge Survey (the source of the data in Table 1, as well as statistics from the American Hospital Association and from Medicare. Each of these sources show that 1975 was not unusual in the regional variation of hospital use.

We sought information on reported morbidity by region for other years to determine whether or not there was something unusual about the regional pattern in 1975. Data from the HIS for selected years 1963 through 1978 are presented in Table 7. The top section provides the age adjusted rates of restricted activity days per person per year by region. Note that the age adjustment used the same four age groups each year, but the standards were based on the U.S. age distribution each year. Hence the total U.S. and age adjusted regional data are not strictly comparable from year to year. However, the data are valid to compare relative morbidity by region in any given year.

The West had the highest number of restricted activity days per person per year for each year, except 1973. The North Central region was more likely to have the lowest number of restricted activity days per person per year than any of the other regions. Thus 1975 was not unusual in the regions' rank order of restricted activity days per person.

The South tended to have the highest percentage of the population with limited activity on an age-adjusted basis, while the Northeast was consistently the lowest by this measure. The data on Table 7 are for the entire HIS population, whereas our study was for those under 65 residing in metropolitan areas. The nonmetropolitan population of the South has a high percentage with limitation of activity. Data for the years 1973 and 1974 combined show that the percentage of the population with limitation of activity residing in metropolitan areas was highest in the West for those under 45 and for those age 45–64.[9] Thus for those two years for which we have data comparable to ours, interregional differences in limitation of activity were similar to 1975.

We do not know whether there are important regional differences in the reporting of morbidity by the HIS for populations that are identical in their "true" morbidity. Given the fact that most in the West are either in-migrants or descended recent in-migrants from other regions of the U.S., we doubt any pervasive cultural difference.

In a recent editorial, Wilson[10] warns that greater use of health care services by respondents can lead them to report more morbidity in the Health Interview Survey. Data in Table 8 show that residents of Western SMSAs did make more visits to physicians per year for each adult age category than those living in other regions of the U.S. We interpret this as a response to the higher morbidity in the West, but Wilson's view cannot be rejected. It may well be that some of the higher physician visit rates in the West are a result of substituting office care for hospital care in that

TABLE 7. Restricted Activity Days Per Person Per Year and Per Cent of Population With Limitation of Activity Selected Years 1963–1978 by Region U.S. Civilian Non-Institutional Population Sample in Health Interview Survey

	Region				
	Northeast	North Central	South	West	U.S.
Restricted activity days per person per year (age adjusted)					
July 1963–June 1965[2]	13.7°*	16.0	17.9^{+}†	17.9^{+}	16.3
1968[3]	17.4	13.8°	16.0	17.8^{+}	15.3
1969 & 1970[4]	14.1	13.0°	15.8	16.4^{+}	14.7
1971[5]	14.6	14.2°	16.7	17.9^{+}	15.7
1972[6]	15.4°	15.7	17.5	18.1^{+}	16.5
1973[7]	13.6°	15.4	18.5^{+}	18.2	16.4
1977[6]	16.8	16.7°	17.9	19.1^{+}	17.4
1978[7]	17.3	16.8°	19.1	20.7^{+}	18.4
Per cent with limitation of activity (age adjusted)					
July 1963–June 1965[2]	9.2°	11.7	14.8^{+}	12.2	12.1
1969 & 1970[4]	10.7°	10.9	13.1^{+}	12.1	11.7
1972[6]	11.4°	12.0	13.5	13.6^{+}	12.6
1973[7]	12.1°	12.6	14.5^{+}	14.1	13.1
1974[8]	13.2°	13.7	14.9^{+}	14.8	14.1
1977[6]	12.0°	12.3	14.0^{+}	13.5	13.0
1978[7]	12.8°	12.8°	14.2	15.0^{+}	13.6

N.B. Includes HIS sample residing in all areas, not just SMSAs.

* ° = Lowest value on each line.

† $^{+}$ = Highest value on each line.

region. The data in Table 8 omit visits to hospital inpatients.

Substantial interregional differences in health status and mortality have been observed by demographers and epidemiologists for many specific conditions.

More research should be done on the nature of the differences in health status by region of the country. Some differences can be attributed to the life-style in the West. For example, the West is high in the number of accidents. But other differences in morbidity may be due to selective in-migration to the West of those suffering from certain chronic conditions such as arthritis and respiratory problems. In general, it has been assumed that in-migrants tend to be healthier than others in the population. While in-migrants are quite young, they may not be healthy.

Restricted activity days are caused by a wide variety of conditions. Obviously some diagnoses are much more likely to be associated with hospital care than other diagnoses (e.g., a serious injury vs. an upper respiratory infection). Some of this variation in causes of restricted activity days is controlled by the use of age-sex specific categories. For example, in 1975 about 30 per cent of all restricted activity days for those under age six were for upper respiratory infections, while this condition comprised under 10 per cent of the total for the entire population.

In the present analysis it has been tacitly assumed that the mix or morbidity for each given age-sex category is relatively uniform in each region of the country, or at least uniform in its need for hospital care. We have done a study (unpublished) for the 1975 HIS West sample similar to the one reported here. In that study we standardized hospital use by 40 diagnosis-specific morbidity measures for six age-sex groups. The expected hospital use from this diagnostic specific approach gave results for four diverse populations that differed only a few per cent from those based on the method used in this analysis. This supports our assumption regarding morbidity mix. However, our approach probably could be improved by giving different weights to restricted activity days with and without bed disability as well as using other measures of the severity of restricted activity day.

Our use of age-specific restricted activity days as a proxy for the types of conditions which lead to hospitalization also is supported by standards shown in Table 3. Apart from infants under age one, the standards had a much narrower range than age-specific hospital use of persons with and without limitation of activity. This

TABLE 8. Annual Physician Visits Per Person by Age and by Region
Residents of SMSAs—1973–1974 Health Interview Surveys*

Age	Region				
	Northeast	North Central	South	West	All U.S.
Under 17	4.6•†	4.3°‡	4.3°	4.3°	4.4
17–44	4.9°	5.1	4.9°	5.8•	5.1
45–64	5.5	5.5	5.4°	6.4•	5.7
65+	7.1	6.6	6.3°	8.0•	7.0
All	5.2	5.2	4.9°	5.7•	5.2

* Adapted from Gentile A. Health characteristics by geographic region, large metropolitan areas, and other places of residence, U.S., 1973–1974. Series 10: Data from National Health Survey, no. 113, (DHEW Publication No. (PHS) 75-1540).

† • = Highest value on each line.

‡ ° = Lowest value on each line.

suggests that much of the variation in hospital use by different populations can be explained by different restricted activity day rates of the populations.

The days spent recuperating at home after a hospital episode are classed as restricted activity days by the HIS. Thus some restricted activity days are a consequence of hospital care. In our model this raises the expected hospital use in areas where hospital use is high. This effect is most prominent for a hospitalized person who would not have been considered restricted in activity by HIS when he entered the hospital (e.g., for some elective surgery, for some diagnostic admissions).

The magnitude of this effect is small. For example in our 1975 SMSA HIS sample the annual hospital nights were 5 per cent of annual restricted activity days. If the average post-hospital convalescence was equal to the time in the hospital then about 10 per cent of restricted activity days could be attributed to either hospitalization or convalescence from hospitalization. The overall effect of counting hospital convalescence as a restricted activity day would be to reduce and not increase differences in morbidity standardized hospital use by region.

Possible Explanations of Regional Differences in Standardized Ratios

There are many possible explanations of the regional differences in standardized hospital-use ratios. Health care coverage is among them. The West was substantially slower than the midwest and Northeast in enrolling its population in Blue Cross hospital insurance after World War II.[11] The states with high Blue Cross enrollment in 1947 tend to be those with longer average lengths of stay in the 1970s. Even today the West is a little lower than the U.S. average in the percentage of its population under age 65 with private health care coverage. This could lower relative hospital use in the region.

The West is the region with the highest percentage of its population enrolled in prepaid group practices. Our analysis of the 1975 HIS showed that 10.4 per cent of the western SMSA population under age 65 were members of prepaid group practice plans (PGP). Over many years the competition between prepaid group practice and other private coverage in the West may have led to less hospital use in that region. Prepaid group practice plans have two effects on utilization. They have less hospital use by their members (relative to health status) and they also may lower hospital use of those with competing private health care coverage within their service areas. The influence of PGP in the West is a key hypothesis to be tested in seeking an explanation of the observed regional difference of hospital use.

There may be relatively more surgery performed on patients in the West who do not spend a night in the hospital even though they use hospital operating and recovery rooms. Unfortunately, there are not good national statistics on the volume or trends of such nonadmitted surgical cases, but they are quite common in Western prepaid group practices.[12] It also is possible that home health care differs by region of the country, although such care generally is considered a substitute for long-term, not short-term, hospital care.

The West has been a leader in developing insurance for office dental care, which can reduce the demand for inpatient dental surgery. The number of inpatient dental surgery operations per hundred thousand in the Northeast region is 293 while it is only 54 in the West.[13] This is one small example of the influence of health insurance coverage on hospital use.

Some of the regional differences in use, particularly for cases that are not surgically treated, may be attributable to regional differences in providing care of mental illness and other long-term hospital care in short-term hospitals. The Health Interview Survey only determines whether or not a hospital is short term and not whether care was in a long-term unit of a short-term hospital.

At one time, the West had relatively fewer medical schools and teaching hospitals than other regions of the country and even today the West has a smaller proportion of its hospital use in large teaching hospitals than other regions. This may contribute to the shorter length of stay in the West since large teaching hospitals generally have higher lengths of stay than other hospitals, even after adjusting for case mix.

The West has not been short of hospital beds and this is not a likely explanation of low hospital use in the region. For example, in 1978 the average occupancy rates for general hospitals in the U.S. were as follows[14]:

Northeast	80 per cent
North Central	74 per cent
South	72 per cent
West	67 per cent

Many planners have been concerned about the low occupancy in the West since they consider it a lure to overuse. Our data demonstrate that it is clearly possible for a region to have low use, even with substantial empty hospital beds. This is not consistent with the view that bed supply is important in determining use, the so-called Roemer's Law. At best, this law is only half true. Obviously an area cannot use hospital beds that it does not have, but an area need not use its available beds.

Some may believe that regional differences in medical education are the source of interregional variation in hospital use, since physicians admit all patients to hospitals. The majority of physicians practicing in the West, however, received their education at medical schools elsewhere, including a high proportion at schools in north central states. It would be interesting to determine how long it takes a physician with a medical education in the midwest to conform to Western hospitalization conventions when entering practice in the West. In fact, a good deal might be learned on regional hospital use differences by

TABLE 9. Estimated Reduction in U.S. Hospital Nights If All Regions Had the Same Indirectly Standardized Hospital Night Ration As the West
Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

		Hospital Nights				
					Annual Use by Sample	
Region	Number of Persons	Annual Expected Rate per 100 Persons	Indirectly Standardized	Annual Actual Rate per 100 Persons	Actual	Expected Indirectly Standardized If Ratio Was .644
	A[*]	B[†]	C[‡]	D[§]	E	F[¶]
Northeast	19,506	68.59	1.094	75.04	14,637	8,616
North central	19,045	63.75	1.276	81.35	15,492	7,819
South	19,158	72.19	0.994	71.76	13,747	8,907
West	14,651	86.76	0.644	55.87	8,186	8,186
All	72,360	71.95	1.000	71.95	52,062	33,528

[*] From line c of Table 4.
[†] From line 1 of Table 4.
[‡] From line j of Table 5.
[§] Column B × Column C.
(Column D × Column A) Y 100.
[¶] (Column A × Column B × 0.644) Y 100.

surveying practice changes of physicians who go from the North Central to the Western states and those with the opposite migration pattern.

Implications

The wide range of indirectly standardized hospital nights in the four regions of the U.S. raises important questions on the appropriateness of hospital care throughout the country. We cannot determine from our data the amount of hospital care that is appropriate for any person. Therefore, we can only identify populations that are relatively high or relatively low in their observed hospital use adjusted for their reported morbidity.

This study raises cause-and-effect issues on the matter of hospital care. For example, are those residing in the North Central region healthier because they have more hospital care? And, are those in the West less healthy because of less hospital care? It is essential to obtain information on this issue before concluding which region has the most appropriate amount of hospital care. We are inclined to believe that hospital overuse is a greater problem than underuse, particularly for the great majority of the urban population under age 65 that have conventional health coverage. When the patient has more comprehensive coverage for care in the hospital than for care outside of the hospital, both the physician and the patient have incentives to prefer the hospital over the office as a place for care.[15]

The diversity of health care services in the U.S. provides many large scale natural experiments in alternative ways to deliver health care. In order to interpret the influence of different health care delivery systems it is necessary to standardize for the health risks in the various study populations. This paper presents one method of standardizing that may help in the interpretation of the results of these natural experiments.

The method of indirect standardization presented in this paper is applicable to the vast store of data in current and past Health Interview Surveys. The HIS is a potential treasure trove for studying the epidemiology of health care in the U.S. despite some limitations (e.g., no data on decedents or the institutionalized). Modern computer programs simplify making analyses of the type presented here and public-use tapes of the Health Interview Survey are readily available at a modest cost to investigators. Further work could and should be done on the epidemiology of hospital use as it relates to health status, other health care services and socioeconomic variables. For example, the present study should be redone for the nine census divisions of the country. It would be possible to provide reasonably accurate data for a half dozen of the largest metropolitan areas. (Analyses of areas smaller than the four major census regions may require pooling of several years' data to augment sample sizes.) We have already compared the central cities with the balance of SMSAs by region of the country. While the central cities are higher in expected hospital use per capita than their surrounding counties, the ratio of actual to expected hospital use within any given region is very similar for the central cities and their adjacent counties.

One implication of our findings is that the opportunities for modifying hospital use through either regulatory means (e.g., rate regulation), voluntary efforts (e.g., hospital review committees) or market competition vary considerably by region. In essence, a region with a low indirectly standardized hospital-use ratio may have less opportunity to lower its use than another area with a high indirectly standardized ratio. Retrospective evaluations of regulatory and competitive experiments should take into account these regional differences in potential for change.

The PSROs (Professional Standards Review Organization) have sought to determine whether observed differences in hospital length of stay for hospitals within a given area are justifiable, but they have tended to use local standards to make this judgment. We believe that the differences in hospital use between various regions are more important. It might be instructive to have physicians from the west make rounds at north central hospitals (or serve on north central PSROs) and vice versa.

Any national health insurance program should consider regional differences in health status when allocating health care resources. Obviously, less healthy populations warrant more care, all other things being equal. Regional differences in health status should be included as a consideration in the monitoring of government programs such as Medicare and Medicaid. The interregional variations in age specific morbidity in Table 2 demonstrate one problem of using age and sex along as substitutes for health status.

We found more interregional variation in nonsurgical hospital case rates than in surgical case rates. Perhaps some reconsideration should be given to the concept of discretionary hospitalization, which some equate with elective surgery. Elective surgery may be discretionary with respect to the time and place of the operation but some elective surgery is not really discretionary in terms of its necessity. For example, the surgical repair of some severe congenital malformations may be quite necessary even if the surgery is elective in time and place. In contrast, many medical admissions to hospitals may be urgent in time but discretionary in terms of the necessity of hospitalization. Criteria for medical admissions are likely to be less rigorous than those for surgical admissions and hence, more subject to nonmedical considerations by both patient and physician.

The general shift of the U.S. population to the South and West continues. If these in-migrants follow the patterns of hospital care in these two regions, then this population shift should reduce overall hospital care for a U.S. population constant morbidity. This is because people are moving from geographic areas with higher indirectly standardized use ratios to those with lower ratios.

There are a few studies of the secular trends in health status in the U.S. A recent article[16] presents some data from the Health Interview Survey between 1966 and 1976 and concludes that more morbidity is being reported by the National Health Survey Samples over time. Our data in Table 7 lead to the same conclusion. Changes in the age-sex specific morbidity of the population over time may be contributing to health care utilization trends. It is quite possible for disability to be increasing, even while death rates decline. Prolonged life could be associated with more limitation of activity and restricted activity days.

It is tempting to speculate on the cost implications of our finding. If one assumes that the volume of hospital care in the West is at least sufficient (considering the morbidity of its population) then the indirectly standardized hospital ratio of 0.65 in the West implies that the U.S., including the West, would use 35 per cent fewer hospital nights if it conformed to Western practices. Because the logic of this estimate was not obvious to several reviewers, Table 9 was prepared to show its basis. The sum of the regional data in column E is the total actual annual hospital nights used by the sample population. The sum of the regions in Column F is the number of nights that would have been used by the study sample if each region had had an indirectly standardized ratio of 0.644 (i.e. the same as the West). The expected use divided by the observed use is 0.644 (i.e. 33,528 ÷ 52,062). Thus, although there would be no change in the West by adopting Western standards, the average reduction in the other three regions would be greater than 35 per cent.

Since hospital days for the less sick are likely to comprise much of the reduction in use, the percentage that hospital inpatient costs were reduced would be less than the percentage that the use was reduced.[17] Nonetheless, the total dollar savings could be 20 to 25 per cent of all inpatient U.S. short stay hospital costs, currently estimated at about $80 billion per year. In addition, there should be some savings in physician and surgeons costs, as well as a reduction in the iatrogenic morbidity which is a common consequence of hospital care.[18, 19]

Acknowledgments

D. Taya Dunn Mangnall performed all of the computer analyses for this study using the computer resources of the Northern California Region of Kaiser Foundation Health Plan, Inc.

References

1. National Center for Health Statistics. Health United States 1980. Hyattsville, Md.: National Center for Health Statistics, 1980 (DHHS Publication No. (PHS) 81-1232).

2. Wilder CS. Health Characteristics by Geographic Region, Large Metropolitan Areas, and Other Places of Residence, U.S. July 1963–June 1965: National Center for Health Statistics, (Series 10: Data from the National Health Survey, no. 36) (PHS) 1000.

3. Wilder MH. Disability Days, United States–1968, (Series 10: Data from the National Health Survey, no. 67) (DHEW Publication No. (HSM) 72-1034).

4. Namey CW, Wilson RW. Health Characteristics by Geographic Region, Large Metropolitan Areas, and Other Places of Residence U.S. 1969-70, (Series 10: Data from the National Health Survey, no. 86) (DHEW Publication No. (HRS) 74-1513).

5. Wilder MH, Pearson AN. Disability Days U.S., 1971 (Series 10: Data from the National Health Survey, no. 90) (DHEW Publication No. (HRA) 74-1517).

6. National Center for Health Statistics, Health United States 1979, Hyattsville, Md.: National Center for Health Statistics, 1979 (DHEW Publication No. (PHS) 80-1232).

7. National Center for Health Statistics. Health United States 1980. Hyattsville, Md.: National Center for Health Statistics, 1980 (DHHS Publication No. (PHS) 81-1232).

8. Wilder CS. Limitation of Activity Due to Chronic Conditions: U.S., 1974, (Series 10: Data from the National Health Survey, no. 111) (DHEW Publication No. (HRA) 77-1537).

9. Gentile A. Health characteristics by geographic region, large metropolitan areas, and other places of residence, U.S., 1973–1974, (Series 10: Data from National Health Survey, no. 113) (DHEW publication no. (PHS) 75-1540).

10. Wilson RW. Do health indicators indicate health? Am J Public Health 1981;71:461.

11. Reed LS, Blue Cross and medical service plans. Washington, D.C.: Federal Security Agency, 1974.

12. Marks SD, Greenlick MR, Hurtado AV, Johnson JD, Henderson J. Ambulatory surgery in an HMO: a study of costs, quality of care and satisfaction. Med Care 1980;18:127.

13. Ranofsky AL. Utilization of short stay hospitals: annual summary for the United States, 1975. National Center for Health Statistics, (Series 13: Data from the National Health Survey, no. 31) (DHEW publication: No. (HRA) 77-1782).

14. Strahan, GW. Inpatient Health Facilities Statistics U.S., 1979. Data from the National Health Survey Series 14, No. 24. (DHHS publication no. (PHS) 81-1819).

15. Blumberg MS. Health status and health care use by type of private health coverage. Milbank Mem Fund Q 1980;58:633.

16. Colvez A, Blanchet M. Disability trends in the U.S. population 1966–1976: analysis of reported causes. Am J Public Health 1981;71:464.

17. Blumberg MS, Gentry DW. Routine hospital charges and intensity of care: a cross-section analysis of fifty states. Inquiry 1978;15:58.

18. Couch NP, Tilney ND, Rayner AA, Moore FD. The high cost of low frequency events—The anatomy and economics of surgical mishaps. N Engl J Med 1981;304:634.

19. Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at the university hospital. N Engl J Med 1981;304:638.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population?

NORALOU P. ROOS, PH.D.,* AND LESLIE L. ROOS, JR., PH.D.†

Major unexplained variations in surgical rates have been documented across geographic areas over at least the last four decades. This paper attempts to assess the contribution that population characteristics make to these variations. Variations in surgical rates among the elderly across 56 small rural areas are compared with variations in characteristics of the elderly population of these areas including self-reported health status, levels of disability, mental status, socioeconomic status, ethnic characteristics, etc. The data do not support a needs model for explaining variations in surgical rates. High rate areas were not characterized by an elderly population that was disabled and in ill health. High rate areas did contain a more educated elderly population, as well as more elderly whose national origins were Canadian, British or American.

MAJOR UNEXPLAINED VARIATIONS in surgical rates have been documented across geographic areas over at least the last four decades. Glover[1] found a tenfold variation in British tonsillectomy rates in 1949, ranging from 40 to 440 per ten thousand children. Vayda and his collaborators noted both the existence of these rate variations across Canadian provinces in 1968 and their persistence over the subsequent ten-year period, the period during which national health insurance was introduced.[2,3] Wide intraprovincial variations in surgical rates within Manitoba,[4] Ontario[5] and Saskatchewan[6] also have been demonstrated.

Variations in surgical rates across small areas within the United States have been found to be as large as, if not larger than, those reported in Canada. Wennberg and Gittelsohn[7] reported that one Vermont hospital service area performed tonsillectomies at a rate of 151 cases per 10,000 residents, compared to the neighboring four hospital service areas whose rates ranged between 32 cases and 39 cases per 10,000 residents. Data from Maine[8] indicate widespread variations in surgical rates for nine different surgical procedures (in-

---

* Social and Preventive Medicine and Business Administration, University of Manitoba, Winnipeg, Manitoba, Canada.

† Business Administration and Social and Preventive Medicine, University of Manitoba, Winnipeg, Manitoba, Canada

Interpretations of the data are the authors' own and do not necessarily represent the official opinion of Health and Welfare, Canada, the Manitoba Health Services Commission or the Manitoba Department of Health.

This research was supported by Grant No. 6607-1157-46 from Research Programs Directorate, Health and Welfare, Canada, by a Career Scientist Award No 607-1001-22 to Dr. Noralou Roos and by a Health Scholar Award No. 606-1114-48 to Dr. Leslie Roos.

Reprint requests: Noralou P. Roos, Department of Social and Preventive Medicine, University of Manitoba, 750 Bannatyne Avenue, Winnipeg, Manitoba R3E OW3, Canada.

0025-7079/82/0900/0945/$01.50 © J. B. Lippincott Co.

cluding appendectomy, prostatectomy, inguinal hernia, etc.), while Lewis[9] found three to fourfold variations in Kansas regional rates for the performance of six common surgical procedures.

Finally, surgical rate variations across countries have been documented. Bunker[10] identified the differences in surgical rates between the United States, England and Wales, while Vayda examined the difference[11] and persistence[12] of differences between Canadian surgical rates and those of England and Wales. McPherson et al.[13] have focused on the patterns of regional variations within and between these three countries.

To date, no one questions the existence of these variations. However, the meaning of these variations is not clear. Most explanations focus upon the link between surgical rates and the availability of physician and hospital resources. Studies have shown "a strong positive correlation between an area's physician-population ratio and the per capita utilization of health services over time."[14] As long as there are more surgeons and more hospital beds available to a population, more surgery will be perfomed. Lewis,[9] in an early study of this relationship, concluded that his results "might be interpreted as supporting a medical variation of Parkinson's Law; patient admissions for surgery expand to fill beds, operating suites and surgeons' time." Although Vayda's earlier work[5] supported this law, more recent analyses of the Canadian data show no consistent relationship between the availability of surgical resources and hospital beds, and the amount of surgery performed.[3, 12]

Bunker[1] was one of the first to suggest that economic factors may help explain the existence of variations in surgical rates. Incentives operating in a fee-for-service system might contribute to variations in rates observed between prepaid and fee-for-service practices in the United States, and to the differences between Britain and the United States.

However, while financial incentives may contribute to surgical rate variation, later research by Bunker and Brown[15] has confused this issue. Elective surgical rates for physicians and their spouses were found to be as high as or higher than such rates in other professional groups. These results suggest that physicians, as informed consumers of medical care, place a high value on surgical treatment.

Despite the uncertainty about the causes of variations in surgical rates, the simple existence of these variations suggests to some that unnecessary surgery must exist. Goran,[16] a former director of the Office of Professional Standards Review Organizations (PSRO), observed the variations in surgical procedures across PSRO geographic areas and argued that the lowest use represents a target level that all areas should strive to achieve.

Others take strong exception to this position.[17] Pauly[18] thinks that significant variations in the prevalence of conditions that prompt surgery would explain part of the variation in rates. Some have expressed the concern that low British surgical rates may be attributed to a failure to meet the existing health needs of the population.[19] Rutkow and Zuidema[20] ask the important question:

> Might not studies comparing case loads in various countries point out socioeconomic differences concerning the delivery of health care within each country as opposed to a true indicator of unjustified surgical rates?

Unfortunately little research has been done on the relationship between surgical rates, and the health and socioeconomic status of a population. Our tonsillectomy study in Manitoba[21] was one of the few efforts attempting to correlate surgical rates with morbidity rates. No relationship was found between the tonsillectomy rate in an area and the prevalence of respiratory morbidity in the relevant population. Conclusions from this paper were necessarily tentative: only nine regions were examined, morbidity data were obtained from claims rather than from patient self-reports and only one surgical procedure was studied.

Wennberg and Fowler[22] have examined the relationship between surgical rates and population characteristics in a different fashion. They identified six geographically contiguous areas in Vermont known for their variations in surgical rates and conducted three hundred household interviews in each area. Differences in population characteristics across these areas did not explain their different use of health care. Illness rates and the extent of insurance coverage varied little; existing income differences were not related to variation in the consumption of health care. While Wennberg and Fowler concluded that consumers probably do not determine variations in rates of health care, this study also was handicapped by the few areas available (six) as well as by the limited indicators of surgery (only overall surgical rate is discussed) and population morbidity.

The present research examines the relationship between surgical rates and population characteristics over 56 hospital areas in rural Manitoba. This paper focuses on surgery in the elderly (those 65 and over). The elderly are an important group in which to study these questions. Certain types of surgery, including lens removal and hip surgery, offer the ability to function independently. Yet, the elderly patient experiences a higher surgical complication rate, slower recovery and higher postoperative mortality. Therefore, both the positive and negative implications of high and low rates become important.

This paper examines overall operative rates among the elderly, as well as the amount of elective surgery (in general and for one specific procedure) and the rate of complex surgery. Variations in surgical rates across the 56 areas are compared with variations in important characteristics of the elderly—self-reported health status, levels of disability, mental status, socioeconomic status, ethnic characteristics and so forth.

This analysis addresses several important questions in the literature. Pauly[18] has put forward a needs model for explaining the variations in surgical rates. The needs model might predict that the worse the overall health of an area's population, the more need for surgical intervention. The higher the illness rate in an area (as measured from interviews with the elderly in 1971), the more surgery expected to take place in that area. Conversely, the healthier the population, the less surgery expected in that area.

In a more specific test of the relationship between health characteristics and subsequent need for surgery, the needs model also would predict that the more area residents reporting "eye trouble not relieved by glasses," the greater the need for corrective surgery, and hence, the higher the lens-removal rate.

The socioeconomic explanation of surgical rate variations[20] generates different hypotheses about the relationship between consumer characteristics and surgical rates. This model suggests that economic differences are a major contributor to the differences in surgical rates between the United States and Britian. The United States' ability and willingness to spend more money on health care supports more surgery, presumably to the benefit of the population.

Canadians enjoy universal health care insurance, with no deductibles that would affect access to surgery (except for cosmetic plastic surgery). Differences in surgical rates can not be attributed to residents' differential abilities to pay for medical care, although an economist might be concerned about income foregone during convalescence. If individuals with higher incomes and higher education show more sophistication in the use of medical care and place a high value on surgery, areas with disproportionate numbers of highly educated, wealthier elderly should show higher surgical rates. This relationship might be particularly strong for elective surgical rates, characterized by their impact on quality of life, not their life-saving potential.

Method

This research analyzes the relationship between the health and other pertinent characteristics of a small area's elderly population and the area's surgical rate. As developed by Wennberg and his collaborators,[7] small area analysis involves developing population units (to obtain the denominator) and counting all health care utilization by this unit regardless of where it takes place. A denominator counting all individuals living in an area makes it possible to adjust the utilization experience of the population by age and sex, removing a possible patient-related contribution to surgical rate variation.

This paper focuses on surgery among the Manitoba elderly resident outside of Winnipeg, the provincial capital. Winnipeg residents, over half the population of the province, are excluded because service areas in a major metropolitan area are more difficult to define. However, as noted above, any rural Manitoban having surgery performed in Winnipeg will have this surgery attributed back to his home area.

A total of 56 hospital areas have been defined. Five of the areas contain two hospitals, while 51 have one hospital. Table 1 summarizes the most important characteristics of the 56 hospital areas. Because of the range in most of the numbers making up the table, reporting the median (the case lying exactly in the middle once all the cases have been rank ordered from highest to lowest), rather than the mean, seemed more appropriate.

The rural hospitals are mainly small (the median size being 24.5 beds) and have low occupancy rates (the median rate being 76 per cent). Rural Manitoba also has an ample supply of hospital beds, the median rate per thousand population falling at 6.3. row D of Table 1 reports the percentage of patients in municipalities assigned to a specific hospital or hospitals that receive their care within the hospital(s) in their area. The area usage figures are somewhat lower than those usually reported by Wennberg, partly because of rural patients' heavy usage of Winnipeg hospitals. (A high percentage of Manitoba's specialists are located in Winnipeg.) The figures in rows D and E were calculated using the entire population of the hospital areas, not just the elderly population.

Row F records the size of the target population in the 56 areas. While the num-

TABLE 1. Characteristics of 56 Hospital Areas

	Median	Minimum	Maximum
Hospital Characteristics			
A. Hospital size (# of Beds)	25.5	7	219
B. Occupancy rates	76.2%	49.3%	93.4%
C. Beds/1000 population	6.3	3.1	23.7
D. Per cent of patients in area who use their area hospital (vs. going elsewhere)	57.3	32.6	85
E. Per cent of patients using area hospitals who are from area	70.9	26.7	100
Size of Target Population in Area			
F. Size of elderly population (66+) in each area	541	141	3713
G. Number of individuals in area interviewed in 1971	37	6	395

bers are not large, all analyses have been run with and without the 11 areas having fewer than 250 elderly. Rates in these areas might be expected to be least stable. Results are reported, where noticeably different.

Results

Interviews With the Manitoba Elderly

In 1971 extensive interviews with a large probability sample of Manitoba's elderly were conducted.[23,24] This survey included residents in the community, hostels, personal care homes, extended care units and mental institutions. Only the very few elderly who were long-term residents of acute hospitals were excluded from the sampling frame.

A 5 per cent probability sample was drawn from persons 65 and over living in rural communities (not in a facility). Those living in facilities (nursing homes, hostels and public housing units) were sampled at a rate of 20 per cent. In calculating the characteristics of the areas from interview data, the numbers have been reweighted to approximate the general population.

Two parallel samples were drawn using the master registry of the Manitoba Health Services Commission. When an individual could not be located or refused to participate, a counterpart from the second sample was randomly assigned to help maintain appropriate sampling for statistical purposes. Seventy-two per cent of those interviewed were drawn from the first sample; 28 per cent of the respondents came from the second sample. No significant differences were found between the two samples along several important demographic (marital status, etc.) and utilization (hospital admissions, etc.) characteristics.[25]

The sample appears to accurately reflect several characteristics (age, sex and marital status) of Manitoba's elderly population. The sample somewhat underrepresents the more educated elderly (33 per cent of the sample, as compared to forty per cent of the province's elderly, report nine or more years of education) and underestimates the rate of hospitalization by approximately 15 per cent. This bias, even in a study using proxy interviews, multiple callbacks and an all-inclusive sampling design, suggests the difficulty of accurately identifying a representative elderly sample. A total of 2,626 elderly were interviewed from our 56 rural hospital areas.

Health Status

The interview itself contained 517 separate measures of personal, social, health and demographic characteristics. This analysis focuses on a small subset of the best-known and validated measures. Six survey measures of health status are presented in the top part of Table 2. For each measure, the mean and standard deviation are reported across the 56 areas. When the smallest areas are excluded, the mean never changes by more than 2 percentage points, although the standard deviation (as one would expect) is reduced by up to 25 per cent. Most of the items are self-explanatory; both self-perceived health status and the number of health problems have been shown to predict health services utilization (number of physician visits and hospitalizations).[25]

The next to last health status measure is the Shanas Disability Index which predated the Katz[26] activities of daily living index; the two contain similar types of items. Both the Katz and the Shanas indices have been criticized for not measuring whether an individual needs help with various activities, but whether the individual is receiving help (needed or not). This may be a particular problem where an elderly individual is living with a spouse. We have attempted to eliminate this bias by recalculating the Disability Index, including only those area respondents living alone. The last health status measure, therefore, reports the per cent of those elderly interviewed in an area who live alone and show

TABLE 2. Additional Characteristics of the Areas' Elderly Population: Health Status, Mental Status, Life Satisfaction, Access to Health Care and Socioeconomic Status

	Mean	Standard Deviation
Health Status		
Per cent of respondents in area reporting that health is good or excellent for their age	59.5	13.71
Per cent of respondents in area reporting none of a series of 15 health problems they were asked about, including such things as stomach trouble, arthritis etc.	12.8	7.69
Per cent of respondents reporting no eye problems	79.4	9.78
Per cent of respondents in area reporting spending no time in bed sick in the last year	83.1	9.85
Per cent of respondents in area showing no functional disability from responses to 13 questions including do you need help with dressing, medication, housework, managing stairs, etc.	80.9	7.97
Per cent of respondents living alone who show no functional disability	84.2	14.94
Life Satisfaction		
Per cent of respondents in area scoring Excellent (17 or higher) on life satisfaction index (20 agree-disagree questions including "I feel old and somewhat tired")	26.6	14.22
Access to Health Care		
Per cent of respondents in area saying no difficulty seeing doctor (or dentist)	89.8	10.88
Per cent of respondents in area who report having regular physician	89.1	8.20
Socioeconomic Status		
Per cent respondents in area with 9 or more years education	27.2	15.41
Per cent of respondents in area who say income and assets currently satisfy their needs very well or adequately	76.0	12.01
Per cent of respondents in area whose income was above official poverty line*	16.3	8.5
Ethnic Status		
Per cent of respondents in area whose national origins were Canadian, American or British	52.8	28.50

* This variable was based only on the responses of those elderly living in the community or in public housing units.

no functional disability. The mean for this measure is somewhat higher than that for the previous measure, suggesting either that disability indices overestimate the disability in individuals not living alone, or that individuals may be able to live alone simply because they are less disabled.

The measures of health status are intercorrelated. Areas with a high percentage of respondents reporting good health were also areas where respondents were more likely to report no health problems (r = 0.44, sig 0.001).[‡] Such areas also were more likely to have respondents report to have spent no time being sick in bed in the last year (r = 0.28, sig 0.02). However, the self-reported health status of respondents in an area and the disability indexes were not related.[§]

Life Satisfaction, Access and Socioeconomic Status

The index of life satisfaction is a thoroughly researched, widely used measure in the elderly population.[28,29] To the degree that elective surgery might be sought by individuals wanting to improve the quality of life, low satisfaction scores might be associated with high surgical rates.

Two measures of access to health care include the percentage of respondents in an area reporting no difficulty in seeing a doctor (or a dentist) and the percentage of respondents in an area having a regular physician. These measures were correlated across the areas (r = 0.33, sig 0.01). The association (or lack thereof) of these variables with surgical rates will help test the Rutkow-Zuidema[20] suggestion that low surgical rates occur where the population has restricted access to medical care.

The three measures of socioeconomic status include the percentage of respondents in an area with nine or more years of education, a subjective measure of whether the respondents' income satisfies their needs, and an objective measure of the percentage of elderly in an area with incomes above the official poverty line. In calculating this latter measure we have excluded individuals who were in nursing homes. Such individuals have their food and shelter provided and have quite different income needs than those living in the community.

As Table 2 indicates, although 76 per cent of respondents across areas reply that their income and assets currently satisfy their needs very well or adequately, only 16 per cent (on average across areas) of the respondents were above the poverty line in 1971. (The 1971 Canadian poverty line figures were $204 a month for single adults, and $375 a month for a family of two.) While these figures reflect the elderly's dismal economic plight, some of the rural elderly live on farms and have income in kind not reflected by these figures.[‖]

The final measure in Table 2 indicates the ethnic heterogeneity of the area. The standard deviation of this measure (the per cent of respondents with Canadian, American or British national origins) reflects Manitoba's ethnically distinct communities. It will be possible to determine if Anglo-Saxons use surgical care differently from other ethnic groups.[¶]

Surgical Rates

Data on surgical rates are taken from the hospital discharge files of the Manitoba

---

[‡] Note, the Pearson correlation coefficient used throughout this paper is based on an N of 56 hospital areas unless otherwise noted. Since we are dealing with population-based surgical rates, not sample data, the level of significance has no universally accepted meaning.[27]

[§] In a separate analysis, an area's overall hospitalization rate as judged from the claims data was strongly associated with the amount of hospitalization reported by those being interviewed in an area. This was a useful check on the validity of the two measures in the different data sources.

[‖] As a partial check, small area census data on education (mean years of education) and income (mean individual and mean household income) for the whole adult population were compared with the survey data from the elderly sample. The correlation between the two data sets was 0.68 for the education variables. The census income variables showed reasonable correlations with the objective measure of poverty among the elderly (0.47 and 0.44), but not with the subjective income measure (0.18 and 0.24). The subjective and objective income measures of income also correlated poorly at the individual level.

[¶] Adult census data also were used as a partial check on this survey measure. The census data on percentage of English-speaking people in each area were highly correlated (0.76) with the survey measure of Canadian, American or British national origins.

Health Services Commission. The whole population of Manitoba is covered by provincially administered health insurance. Hospital and medical care are documented in considerable detail because of the fee-for-service payment arrangement for medical services.

The data base has been fully described elsewhere.[30,31] The quality of data taken from health insurance claims has been the focus of intensive investigation. The data appear to be an accurate, reliable and valid representation of events in the health care system. In tonsillectomy research, claims diagnoses were logically consistent with surgical procedures subsequently or previously performed. More recently, the claims filed by surgeons for operative procedures have been compared with the hospital discharge claims filed independently by hospitals. A high level of correspondence between information contained in the two files was found. Over 90 per cent of the claims filed by surgeons for hysterectomies were confirmed (by date and type of procedure) in the hospital file, and vice versa.

While the interviews are based upon a sample of the elderly interviewed in 1971, the rate data encompass the entire surgical experience of the areas' elderly population. To generate more stable surgical rates by increasing the size of the numerator, two years of surgery (1971–1972) were combined; an average surgical rate/1000 elderly then was calculated. Moving from rates based on one year's data to those produced from two years' data improves statistical stability; the standard deviation is reduced from 23.6 to 16.6 and correlations with subsequent periods were increased. The rates for 1971–1972 were correlated 0.38 with 1974–1975 rates using the Pearson correlation coefficient (significant at 0.002); excluding the 11 smallest areas with 250 or fewer elderly, the coefficient is 0.54 ($p = 0.0001$). Four years' data (1970–1973) on cataract surgery were used. The cataract surgery rates for 1970–1973 were correlated 0.45 with 1974–1977 rates ($p = 0.0001$); with the smallest areas excluded, the correlation was 0.43 ($p = 0.01$). While larger areas would produce more statistically stable rates, making more than one year's data unnecessary, the smaller areas used here are more likely to capture important practice pattern and population differences.

A surgical discharge was defined as any discharge that included one or more "treatment" surgical procedures (a numeric three-digit International Classification of Diseases Adapted—8 surgical code). The rates calculated for each area included the number of individuals coming to surgery, the number of discharges containing one or more surgical procedures and the hernia equivalent (H.E) surgical rate, a weighted measure based upon the California Index of Relative Values developed by E.F.X. Hughes.[32] A fourth measure, the number of discharges with surgical procedures having a hernia equivalent of 0.5 or greater, is used most frequently in this report. This cutting point excludes such minor surgery as D & C procedures (H. E. weight of 0.43), myringotomies (H.E. 0.09), etc., but includes more complex procedures such as appendectomy (H.E. 1.14), inguinal hernia repair (H.E. 1.0) and so forth. The four variables were intercorrelated 0.75 (sig 0.001) or higher. All surgical treatment, whether performed on an inpatient or outpatient basis, was available for this analysis. However, since on average outpatient surgery only contributed 6 per cent to an area's total complexity score, analysis has been restricted to inpatient surgery.

Table 3 reports the range of variation in the measures of surgical treatment. All rates have been adjusted by age and sex, using the indirect method. The first four measures summarize the amount of surgery in each of the 56 areas; rates varied 2.4- to 2.8-fold across the areas.

The amount of elective surgery also differed from area to area: from 17.6 to 56.5

TABLE 3. Surgical Utilization Characteristics of Hospital Areas

	Median	Minimum	Maximum
All Surgical Treatment			
Number of individuals coming to surgery	85.1	53.3 (60.3)*	144.2
Number of discharges with one or more surgical procedures	96.5	67.8 (69.7)	164.0
The hernia equivalent surgical rate (total complexity)	133.2	82.1 (98.7)	225.9
Number discharges with surgical procedures having H.E. GE 0.5	75.7	52.0	124.0
Elective Surgical Rate †	32.6	17.6 (20.6)	56.5
Rate of Performing Complex Surgery (H.E. of 2.0 or greater)	32.2	14.3 (20.4)	53.5
Cataract Surgery Rate ‡	7.2	3.1	13.1

* When excluding the areas with 250 or fewer elderly changes the minimum or maximum, the new figures are listed in parentheses.

† Elective surgery includes the following procedures: Lens extractions (14.1–14.5); prostatectomy (58.1–58.3); hysterectomy (69.1–69.4, 70.1–70.2); cholecystectomy (43.5); hemorrhoidectomy (51.3); varicose vein stripping (24.4); inguinal hernia repair (38.2). All numbers in parentheses refer to ICDA-8.

‡ Annual rate is based on average of 4 years' observations, 1970–1973.

per 1,000 elderly. Seven of the procedures most commonly referred to as discretionary or elective are included in this variable. By restricting the analysis to the procedures having a hernia equivalent of 2.0 or greater, the complex surgery measure counts the following kinds of major surgery; vagotomy and pyloroplasty (H.E. = 2.00), bilateral nephrectomy (4.29), bilateral orchiopexy (2.14), bilateral fixation of testis (2.4), etc.

The final measure records the variation in cataract surgery across the 56 areas. Over the four-year period examined, residents of high rate areas were on the average almost four times as likely to receive this procedure as were residents of low rate areas.

Table 3 showed the range of variation in, and the rate of, the elderly's exposure to different types of surgery. Table 4 outlines the wide scope of surgical practice patterns from a different perspective. Here the areas are characterized according to the tendency to experience one type of surgery versus another. While in Table 3 the elderly population was the denominator for the rate calculations, in Table 4 the number of 1971–1972 surgical procedures

TABLE 4. Patterns of Surgical Practice Percentage of All Surgery (H.E. 0.5 or greater) in Area That Is of the Specified Type

	Median	Minimum	Maximum
Per cent of all surgery that is:			
Elective*	40.3	28	68 (62)†
Complex (H.E. of 2.0 or greater)	40.3	27 (32)	58 (55)

* All variables are defined as in Table 3.

† When excluding the areas with 250 or fewer elderly changes the minimum or maximum, the new figures are in parentheses.

with a hernia equivalent 0.5 or greater is the denominator. The first row suggests that in at least one area, 68 per cent of all surgery (with a hernia equivalent score of 0.5 or more) consisted of one of the seven elective procedures; in another area, only 28 per cent of the surgery was elective.

Surgical Rates and Population Characteristics

Pearson product moment correlations between the characteristics of the elderly interviewed in an area in 1971 and the area's 1971–1972 surgical rates were examined. The data do not support a needs model for explaining variations in surgical rates. High rate areas were not characterized by an elderly population that was disabled and in ill health. In fact, no consistent stable relationships between the health status variables and the surgical rate variables were found. Those statistically significant relationships that did appear (between surgery and no time sick in bed) vanished when the smallest areas were eliminated. The number of elderly reporting eye problems in an area was also unrelated to the frequency of cataract surgery. Grouping the areas according to 1971–1972 surgical rates, as is done in Table 5, further indicates weak and inconsistent relationships.

Correlations between our best quality of life indicator, the widely used life satisfaction index, and the surgical measures, including the elective surgical rate, generally were not statistically significant. This lack of relationship is important, because the major benefit derived from several elective procedures is thought to be quality of life; benefits in terms of years of life saved or cost reduction are difficult to document.[33]

Subjectively perceived access also was unrelated to surgical rates. Some remote rural areas (in 1974 there was 1,105 population per physician in rural Manitoba[34]) may experience a shortage of physicians, but the substantial amount of surgery done in Winnipeg on rural residents makes it difficult to conceive of access in terms of local physician supply alone. Since very few elderly reported problems seeing a physician and most had a regular doctor, barriers to access are unlikely explanations for low surgical rates.

Although neither of the income variables was significantly associated with the surgical rate measures, Table 6 shows that education (the percentage of elderly respondents with nine years or more education) and ethnicity (the percentage of respondents with Canadian, American or British national origins) were strongly and consistently associated with the measures of overall surgical rate.# In those areas with the highest surgical rates, 38 per cent of the respondents claimed nine or more years of education; in the other areas, 26 per cent or less were so educated (Table 5). The ethnic differences vary even more sharply with overall surgical rates. In the lowest rate areas only 41 per cent of the respondents noted their national origins as Canadian, British or American; in the high rate areas, 67 per cent reported such backgrounds. However, neither ethnicity nor education was associated with the two more specific surgical measures: rate of elective surgery and rate of cataract surgery.

Discussion

Either the health characteristics of an area's population are not related to the type and amount of surgery performed on residents, or the measures used for this study are too insensitive to pick up this relationship. Since our indices have been accepted widely for measuring health status and quality of life among the elderly, the latter explanation is unlikely. Even the quite specific data on problems with vision

---

# Generally similar, but slightly weaker, correlations were found using the census data on education, income and ethnicity.

TABLE 5. Illustrating the Relationship Between 1971 Characteristics of an Area Population and 1971–1972 Surgical Rates

Health Status	Areas Grouped According to 1971–1972 Surgical Rates*			
	Areas With Lowest Rates	Medium Low	Higher	Areas With Highest Rates
Health Status				
Mean per cent of respondents in areas reporting their health is good or excellent for their age	59	55	62	65
Mean per cent of respondents in areas reporting none of a series of 15 health problems they were asked about including such things as stomach trouble, arthritis, etc.	12	12	14	14
Mean per cent of respondents in areas reporting spending no time in bed sick in the last year	81	81	85	88
Mean per cent of respondents in areas showing no functional disability from responses to 13 questions including do you need help with dressing, medication, housework, managing stairs, etc.	81	80	84	79
Mean per cent of respondents living alone who show no functional disability	83	84	83	88
Life Satisfaction				
Mean per cent of respoadents in areas scoring excellent (17 or higher) on life satisfaction index (20 agree-disagree questions including "I feel old and somewhat tired")	25	24	29	30
Access to Health Care				
Mean per cent of respondents in areas saying no difficulty seeing doctor (or dentist)	88	89	92	93
Mean per cent of respondents in areas who report having regular physician	88	87	90	93
Socioeconomic Status				
Mean per cent of respondents in areas with 9 years or more of education	24	26	26	38
Mean per cent of respondents in areas who say income and assets currently satisfy their needs very well or adequately	75	74	78	77
Mean per cent of respondents in areas whose income was above official poverty line	17	18	15	16
Ethnic Status				
Mean per cent of respondents in areas whose national origins were Canadian, American or British	41	47	63	67
Number of Elderly Interviewed	671	594	576	785
Number of Elderly in the Population	10538	10048	9372	9328

* Areas grouped according to their age and sex adjusted rate of the number of surgical discharges with hernia equivalent score greater or equal to 0.5.

TABLE 6. Significant Correlations Between Characteristics of an Area's Elderly Population and the Area's Surgical Rates (Age and Sex Adjusted per 1000 Elderly)

	Individuals Having Surgery	All Surgical Discharges	H.E. Surgical Rate	Discharges With H.E. 0.5 or More	Elective Surgery	Complex Surgery	Cataract Surgery
Socioeconomic Status							
Per cent of respondents in area with 9 or more years of education	0.45§	0.43§	0.34†	0.32§		0.28*	
Per cent of respondents in area who say income and assets currently satisfy their needs very well or adequately							−0.25*
Per cent of respondents in area whose income was above official poverty line							
Ethnicity							
Per cent of respondents in area whose national origins were Canadian, American, or British	0.43§	0.40‡	0.26*	0.23*		0.24*	

Levels of Significance of Pearson R:
* = .05.
† = .01.
‡ = .001.
§ = .0001.

showed no systematic relationship to fourfold variations in cataract surgery rates. The data here accord with several observations by Fuchs.[35] In industrialized societies, the relationship between medical inputs, such as surgery, and presumed outputs, such as health, is tenuous. Although we have not measured health manpower directly, the findings seem in line with Newhouse and Friedlander's[36] results showing no systematic link between the amount of health resources and a number of medical indicators of ill health. Given these data, Pauly's[18] supposition that "there may be significant variations in the (unknown) prevalence of conditions that prompt surgery which can explain part of the wide variation in surgery rates" seems unlikely. Although Barker[37] has clearly documented geographic variations in disease in Britain, he also suggests that such variations are more likely to be reflected in autopsy than in surgical rate statistics.

The relationships among income, educational status and surgical rates are less in accord with the literature. The American study of Bombardier et al.[38] previously found income to be positively associated with surgery, although, other things being equal, the more highly educated tended to have fewer operations. Our differences may be due to the nature of the health care system. Income seems a less important variable in Canada, given the insured care available under National Health Insurance. Bombardier et al. also reported the negative correlation between education and surgery to be least apparent in the oldest and the youngest in their sample. Finally, since in our sample education and ethnicity are highly intercorrelated, the relationship between education and surgery may be spurious.

The correlation between the ethnic makeup of a small area's population and surgical rates is an interesting one. Our data suggest that areas with high surgical rates are also those likely to have disproportionately large numbers of elderly of Anglo-Saxon (English-Canadian or American) parentage. Whether such individuals have a cultural bias towards surgery or are more sophisticated consumers of surgical services (the Bunker-Brown model) can not be determined from these data. In fact, a quite different causal mechanism may be operating. Perhaps North American trained physicians have a surgical bias and are more likely to settle in rural areas having a congenial ethnic makeup. Our ongoing study will investigate the relationship among supply variables, surgical rates and mortality rates. In any case, as Bunker and Brown[15] have observed, supply variables and population characteristics would seem more likely to influence "demand for surgery" than be related to "need for surgery."

The importance of being able to distinguish between demand for surgery (whether influenced by patients or physicians) and the "need" for surgery is clear. High surgical rates have clear short-term cost implications. Wennberg[39] has argued that a high-surgery-use strategy (based on costs in high rate areas) adopted across the U.S. population would cost $6.8 billion per year, although a low-use strategy would cost only $3.0 billion (in 1973 dollars). Since the high-use strategy also may have implications for mortality rates,[40, 41] the importance of establishing the efficiacy of this strategy (or lack thereof) in terms of quality of life is self-evident.

Acknowledgments

The author would like to thank the Manitoba Health Services Commission for facilitating access to claims data and to thank Betty Havens of the Department of Health and Social Development for providing access to the Manitoba Needs Survey. John Bunker, Jack Wennberg and Benjamin Barnes provided valuable suggestions concerning this research.

References

1. Glover JA. Tonsillectomy in the school medical service: increased incidence in 1948. Monthly Bulletin Ministry of Health, London 9:62.

2. Vayda E, Anderson GD. A comparison of provincial surgical rates, 1968. Can J Surg 1975;16:18.

3. Mindell WR, Vayda E, Cardillo B. A ten year study of surgery in Canada 1968–1977. Can Med Assoc J 1982. (in press)

4. White Paper on Health Policy, Winnipeg, Manitoba Department of Health and Social Development: Government of Manitoba, 1972.

5. Stockwell H, Vayda E. Variations in surgery in Ontario. Med Care 1979;17:390.

6. Dyck FJ, Murphy FA, Murphy JK, et al. Effect of surveillance on the number of hysterectomies in the province of Saskatchewan. N Engl J Med 1977; 296:1326.

7. Wennberg JE, Gittelsohn AM. Small area variations in health care delivery. Science 1973;182:14.

8. Wennberg JE, Gittelsohn AM. Health care delivery in Maine I: patterns of use of common surgical procedures. J Maine Med Assoc 1975;66:5.

9. Lewis CE. Variations in the incidence of surgery. N Engl J Med 1969;281:880.

10. Bunker JP. Surgical manpower: a comparison of operations and surgeons in the United States and in England and Wales. N Engl J Med 1970;283:135.

11. Vayda E. Comparison of surgical rates in Canada and in England and Wales. N Engl J Med 1973;289:1224.

12. Vayda E, Mindell WR, Rutkow IM. A decade of surgery in Canada, England and Wales and the United States. Arch Surg 1982;117:846.

13. McPherson K, Strong PM, Epstein A, Jones L. Regional variations in the use of common surgical procedures; within and between England and Wales, Canada and the United States of America. Soc Sci Med 1981;273:15A.

14. Reinhardt UE. The physician as generator of health care costs. Health care and the American economy: issues and forecasts, 1978. Chicago, Health Services Foundation, 1979.

15. Bunker JP, Brown BW, Jr. The physician-patient as an informed consumer of surgical services. N Engl J Med 174;290:19.

16. Goran M. The evolution of the PSRO hospital review system. Med Care 1979;(17 supplement:5).

17. Moore FD. Small area variations in health care delivery; a critique. J Maine Med Assoc 1977;68:49.

18. Pauly MV. What is unnecessary surgery? Milbank Mem Fund Q 1979;57:1.

19. Pearson RJC, Smedby B, Berfenstam R, Logan RFL, Burgess AM, Peterson OL. Hospital caseloads in Liverpool, New England and Uppsala Lancet 1968;2:559.

20. Rutkow IM, Zuidema GD. Unnecessary surgery: an update. Surgery 1978;84:671.

21. Roos NP, Roos LL Jr, Henteleff PD. Elective surgical rates—do high rates mean lower standards? N Engl J Med 1977;297:360.

22. Wennberg JE, Fowler FJ, Jr. A test of consumer contribution to small area variations in health care delivery. J Maine Med Assoc 1977;68:275.

23. Mossey JM, Havens B, Roos NP, Shapiro E. The Manitoba longitudinal study on aging. Description and methods. Gerontologist 1981:21:551.

24. Manitoba Department of Health and Social Development. Aging in Manitoba, vol 1–9. Winnipeg, Manitoba Department of Health and Social Development 1973–1977.

25. Roos NP, Shapiro E. The Manitoba longitudinal study on aging: preliminary findings on health care utilization by the elderly. Med Care 1981;19:644.

26. Katz S, Ford AB, Downs TD, Adams M, Rusby DI. Effects of continued care: a study of chronic illness in the home. U.S. Department of Health, Education, and Welfare, December 1972. (DHEW publication. (HSM)73-3010.)

27. Winch RF, Campbell DT. Proof? No! Evidence? Yes! The significance of tests of significance. Am Sociol Rev 1969;4:140.

28. Berkman L. Measurement of mental health in a general population survey. Am J Epidemiol 1971;2:105.

29. Adams D. Correlates of satisfaction among the elderly. Gerontologist 1971;11:64.

30. Roos NP, Henteleff PD, Roos LL, Jr. A new audit procedure applied to an old question: is the frequency fo T and A justified? Med Care 1977;15:1.

31. Roos LL, Jr, Roos NP, Nicol P, Johnson C. Using administrative data banks for research and evaluation. Eval Q 1979;3:236.

32. Hughes EFX, Fuchs VR, Jacoby JE, Lewit EM. Surgical work loads in a community practice. Surgery 1972;71:315.

33. Bunker JP, Barnes BA, Mosteller F. Costs, risks and benefits of surgery. Oxford University Press, New York 1977.

34. Roos NP, Gaumont M, Horne JM. The impact of the physician surplus on the distribution of physicians across Canada. Canadian Public Policy 1976;2:169.

35. Fuchs VR. Economics, health, and post-industrial society. Milbank Mem Fund Q 1979; 59:153.

36. Newhouse JP, Friedlander MJ. The relationship between medical resources and measures of health: some additional evidence. Rand document R-2066-HEW. Santa Monica, Calif. Rand, 1977.

37. Barker DJP. Geographical variations in disease in Britain. Br Med J 1981;283:398.

38. Bombardier C, Fuchs VR, Lillard LA, Warner KE. Socioeconomic factors affecting the utilization of surgical operations. N Engl J Med 1977;297:699.

39. Wennberg JE. Factors governing utilization of hospital services. Hos Prac 1979;14.

40. Wennberg JE, Bunker JP, Barne B. The need for assessing the outcome of common medical practice. Annu Rev Public Health 1980;1:277.

41. Roos NP, Roos LL, Jr. High and low surgical rates: risk factors for area residents. Am J Public Health 1981;71:591.

C5170-1

Medical Care

October 1982
Vol. 20 No. 10
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Quality, Cost, and Health: An Integrative Model

Physician Practice Profiles for CME: Four Investigations Into Sampling Methods

Improving Compliance Among Hypertensives: A Triage Criterion With Cost-Benefit Implications

Social Networks, Beliefs, and the Decision to Seek Medical Care: An Analysis of Congruent and Incongruent Patterns

Assessing the Effectiveness of Neonatal Intensive Care

Improving Physician Compliance With Preventive Medicine Guidelines

A Description of a Means of Improving Ambulatory Care in a Large Municipal Teaching Hospital: A New Role for Nurse Practitioners

PUBLISHED BY
J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Marion Athearn
Jennifer Gardon
Editorial Assistants

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year for individuals, $48.00, for institutions, $55.00; special price U.S. students, $35.00; all other countries except Japan, one year, for individuals, $59.00, for institutions, $68.00. Single copies $6.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. **Copyright © 1982 by J. B. Lippincott Company.** **Second class postage paid at Hagerstown, MD, and additional mailing offices.**

October 1982
Volume 20, Number 10

CONTENTS

ORIGINAL ARTICLES

Quality, Cost, and Health: An Integrative Model. *Avedis Donabedian, John R. C. Wheeler and Leon Wyszewianski* 975

Physician Practice Profiles for CME: Four Investigations Into Sampling Methods. *Lynn Curry and Wayne Putnam* 993

Improving Compliance Among Hypertensives: A Triage Criterion With Cost-Benefit Implications. *Steven R. Eastaugh and Myron E. Hatcher* 1001

Social Networks, Beliefs, and the Decision to Seek Medical Care: An Analysis of Congruent and Incongruent Patterns. *Emil Berkanovic and Carol Telesky* 1018

Assessing the Effectiveness of Neonatal Intensive Care. *Peter P. Budetti and Peggy McManus* 1027

Improving Physician Compliance With Preventive Medicine Guidelines. *David I. Cohen, Benjamin Littenberg, Cheryl Wetzel and Duncan vB. Neuhauser* 1040

A Description of a Means of Improving Ambulatory Care in a Large Municipal Teaching Hospital: A New Role for Nurse Practitioners. *Daniel M. Becker, Arthur M. Fournier and Laurence B. Gardner* 1046

COMMUNICATION

Spending Time With Patients: The Impact of Organizational Structure on Medical Practice. *Fredric D. Wolinsky and William D. Marder* 1051

BOOK REVIEWS

Learning to Live With Angina, by Cynthia B. Wong. *Thomas S. Inui, Benjamin A. Lipsky and Walter J. Kupersmith* 1060

Response to Inui, Lipsky and Kupersmith. *Cynthia B. Wong and Judith C. Watkins* 1061

Response to Wong and Watkins. *Thomas S. Inui* 1062

The Incidence and Economic Costs of Major Health Impairments: A Comparative Analysis of Cancer, Motor Vehicle Injuries, Coronary Heart Disease, and Stroke, by Nelson Hartunian, Charles N Smart and Mark S. Thompson. *David C. Warner* 1062

Medical Economics and Health Finance, by Steven R. Eastaugh. *S. E. Berki* 1064

Rural Medicine: Obstacles and Solutions for Self-Sufficiency, by Stanley S. Wallack and Sandra E. Kretz. *Roger A. Rosenblatt* 1064

PUBLICATIONS RECEIVED 1067

MEDICAL CARE
October 1982, Vol. XX, No. 10

Improving Compliance Among Hypertensives: A Triage Criterion With Cost-Benefit Implications

STEVEN R. EASTAUGH, SC.D.,* AND MYRON E. HATCHER, SC.D.†

Triage method based upon individual patient characteristics is developed to determine the most effective health education interventions for hypertensive patients. The sample included 402 patients randomly assigned to experimental and control groups. The efficacy of a triage process, whereby patients are subdivided into groups that are more predisposed to benefit from a given health education approach, is discussed. From a limited cost-benefit evaluation, it appears that the benefits of the triage method for achieving medication compliance clearly outweigh its costs only in the case of the highly depressed patients (24 per cent of the sample). Different levels of depression were found to be differentially responsive to different health education treatments for achieving adherence to medication. The most depressed patients were more responsive to family member reinforcement, a 156 per cent increase in the number of high compliers as compared to no intervention.

MUCH INTEREST EXISTS today in targeting public health programs to yield the maximum return on investment. The general objective of this study is to develop patient-specific decision rules that serve as a basis for selecting the health education program most likely to encourage adherence to antihypertensive medication. Decision rules are formed by empirically studying the relationships between salient patient characteristics and treatment effectiveness. A specific objective of the design is to measure the potential importance of mental health variables, such as depression, in explaining the effectiveness of health education intervention efforts. The basic research question is whether interaction exists between the effectiveness of different intervention programs and the characteristics of patients participating in the interventions. If there is a significant link between characteristics and effectiveness, then maximum benefits might be achieved if patients are triaged into the high-yield intervention strategies.

There are several reasons for the current impetus to study efficiency in health education programs. The first reason is the

---

* Assistant Professor of Health Economics, Sloan Program of Hospital and Health Services Administration, Graduate School of Business and Public Administration, Cornell University, Ithaca, New York.

† Director and Senior Associate of Health Systems and Resource Planning Associates, Inc., 16026 Burton Lane, Laurel, Maryland.

Supported by National Institutes of Health Grants HL-17016, MH-14567, and 1-32(HL)-07180.

Presented in part to the Operations Research Society of America and the Institute of Management Sciences Joint National Meeting in Washington, D.C., May 7, 1980.

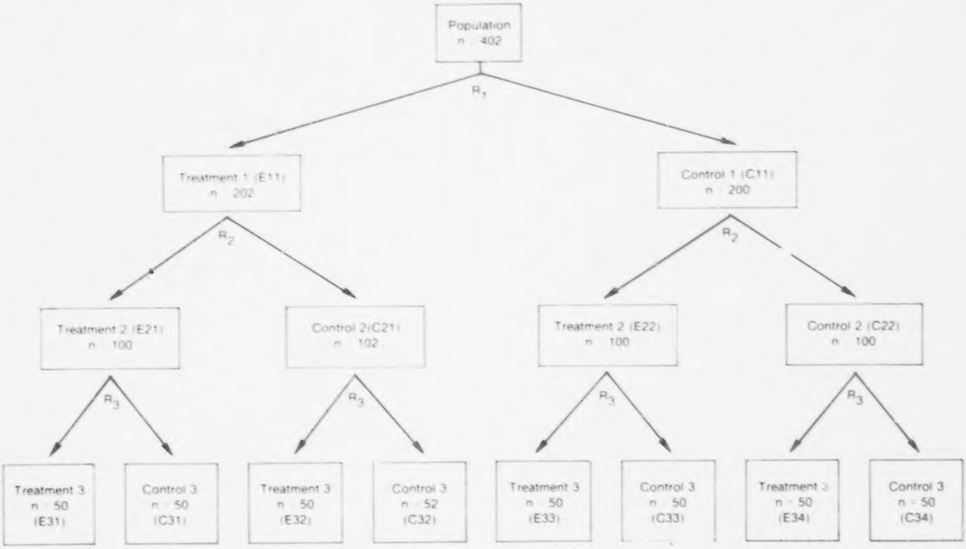
Reprint requests: Steven R. Eastaugh, Sc.D., Assistant Professor of Health Economics, Sloan Program of Hospital and Health Services Administration, Graduate School of Business and Public Administration, Malott Hall, Cornell University, Ithaca, New York 14853.

0025-7079/82/1000/1001/$01.65 © J. B. Lippincott Co.

large body of literature that cites many health education approaches shown to be effective in goal attainment,[1] but unproven from a cost-benefit standpoint. Second, the public cry for efficiency and the demand to conserve resources will necessitate methods of assigning patients to treatments most appropriate to their needs and to their capacities to benefit from them.[2] Third, as is the case with the control of all chronic diseases, adherence to medical regimen through voluntary education is considered a major problem.

The data base consisted of 402 hypertensive outpatients from Johns Hopkins Hospital, who were assigned randomly into three health education interventions: 1) a clarification of the doctor's instructions; 2) family member reinforcement; and 3) a series of three group meetings with other patients. The interventions were applied individually and in sequential combination (Table 1, Fig. 1). Data for application of the methods was derived from a postintervention field questionnaire.

The rationale for all three social support interventions that improve compliance has been reviewed elsewhere.[3,4] Some supporting evidence for the third intervention, group meetings designed to shift the patient's orientation, has been provided by Conte et al.[5] The rationale for the second intervention, family member reinforcement, has been provided by Fass et al.[6] and Levine et al.[7] Less costly forms of intervention, such as the provision of written information to patients, have proven to be ineffective in improving compliance.[8] Physician-directed programs to indirectly improve patient compliance by means of clinician education, even when linked with computer feedback, have proven to be equally disappointing.[9] Although there is an imperfect link between changes in physician attitude and changes in behavior, there is some evidence to suggest


Population
n = 402
R_1
Treatment 1 (E11)
n = 202
Control 1 (C11)
n = 200
R_2
R_2
Treatment 2 (E21)
n = 100
Control 2 (C21)
n = 102
Treatment 2 (E22)
n = 100
Control 2 (C22)
n = 100
R_3
R_3
R_3
R_3
Treatment 3
n = 50
(E31)
Control 3
n = 50
(C31)
Treatment 3
n = 50
(E32)
Control 3
n = 52
(C32)
Treatment 3
n = 50
(E33)
Control 3
n = 50
(C33)
Treatment 3
n = 50
(E34)
Control 3
n = 50
(C34)

R_1, R_2 = Random assignment

R_3 = Stratified random assignment, i.e.—stratified on prior treatment

n = Sample size assigned

FIG. 1. Study Design: sequential factorial provision of three experimental interventions.

TABLE 1. Eight Study Groupings and Percentage of Patients with Blood Pressure Under Control

Study Status	Combinations Utilized*	Sample Size	Percentage of Sample With BP Under Control	
			Initial Baseline Measures	Posttest Following Interventions
E31	RD, FR, SG	50	38%	66%
C31	RD, FR	50	45%	55%
E32	RD, SG	50	40%	53%
C32	RD	52	40%	41%
E33	FR, SG	50	41%	54%
C33	FR	50	37%	48%
E34	SG	50	35%	52%
C34	None	50	41%	42%

* RD is intervention 1 (reinforcement of doctor's message at the exit interview), FR is intervention 2 (family, reinforcement therapy), and SG is intervention 3 (small group therapy).

that attitudes might be instilled more effectively in the last 2 years of medical school.[10]

Health education can be described as a voluntary method for producing change in individuals so that desired behavioral and, presumably, physiologic changes are accomplished. One important principle in health education is the need for continuous application of treatments for any long-term behavioral change.[11] The primary difficulty is the uncertainty of effectiveness and the decay of effect over time. Because of the tendency for a single intervention to decay, health education needs to be a continuous process. Green[12] and Kasl[13] have outlined a number of health education models and treatment interventions designed to affect predisposing, enabling, and reinforcing factors. Unfortunately, many health education efforts have proven to be ineffective.[14]

Background

Hypertension is a largely asymptomatic disorder that individuals tend to disregard and thus not treat. Its estimated prevalence for adults aged 18–79 ranges from 15 to 18 per cent.[15] The prevalence rates are twice as high for blacks as whites and the disease tends to be more severe for blacks. The probability of having hypertension also increases with age. For example about 4 per cent of all individuals in the age groups 17–24 years have hypertension and 35 per cent of all individuals in the age group 65 years of age and over have hypertension. An estimated 9 million Americans have used antihypertensive medication for the condition in the previous 6 months, but 60 per cent of these patients still have high blood pressure and definite hypertension. In summary, about 5 million Americans potentially could be aided by measures that improve the rate of adherence to medication.[16]

For a long time hypertensive disease has been recognized as a serious public health problem that causes morbidity, disability, and mortality. A dilemma of treatment is that a majority of people who have hypertensive disease are symptom free. Even a slight elevation in blood pressure levels causes a marked increase in the risk of major cardiovascular complications. For example, the Society of Actuaries 1959 study showed that 35-year-old men with blood pressure readings higher than 140/90 mm Hg, without concomitant impairments had a 19.4 per cent mortality rate over the next 20 years compared with a 11.0 per cent rate for nonhypertensives.[17] For all causes of deaths, the ten-year mortality rate for men with diastolic pressure 95 to 105 mm Hg was 60 per cent higher than the rate for normotensives. The mortality rate was 200 per cent higher for those with diastolic blood pressure of 105 mm Hg and greater.

Many studies have demonstrated the effectiveness of antihypertension medication.[17,18] The most publicized studies were based on data from the Veterans Administration Cooperative Study Group on Antihypertensive Agents.[20] For a carefully selected population, this prospective study of drug-treated and control-group hypertensives utilizing a double-blind randomized protocol showed that medications will reduce the probability of morbidity and mortality. Increased benefits were a function of the following risk factors: age, race, sex, target organ damage, family history, hyperglycemia, and hypercholesteremia. The 210 VA study patients with moderate hypertension (105–114 mm Hg) had a 3:1 difference in major complications between the treatment and control groups during the 3-year follow-up period. The treatment was efficacious against stroke, congestive heart failure, accelerated hypertension, and progressive elevation of blood pressure, but not against myocardial infarction.[21]

For the VA patients with moderately severe and severe hypertension, 21 out of 70 control patients developed major complications during 18 months of follow-up, but only 1 out of 73 treated patients developed a complication. The treatment given consisted of hydrochlorothiazide, reserpine, and hydralazine hydrochloride. After 2.4 months of active treatment, an average reduction of 42 mm Hg systolic and 29.7 mm Hg diastolic occurred for moderately severe and severe cases. The placebo treatment group showed no significant changes in average blood pressure levels.

At the present time improved adherence to treatment is a greater policy problem than the need to screen populations.[18,22] Many individuals are aware that they have hypertensive disease and have started treatment, but either do not comply to medical instructions or drop out of treatment once they have started. It would not be benefical to screen these patients again.[22] The cost-effectiveness of a nationwide screening program needs careful scrutiny. In this context, Junod and Gutzwiller[23] have designed a medical and public health program evaluation project to pinpoint the most cost-effective modes of reducing cardiovascular disease risk factors. A major component of their cost-effectiveness model is based upon the adherence to medication rate for detected patients.

According to the literature, education, income, occupation, and race are significant socioeconomic factors in determining a person's ability to follow an antihypertensive program.[24] Other characteristics, such as Wallston's scale for locus of control, also have been identified as significant factors that have an impact on the treatment of hypertensives. Naditch found that a perceived external locus of control and a high level of discontent is directly related to having hypertension.[25,25a]

Mental health variables relating compliance behavior or receptivity to health education interventions have not been studied extensively. Bloom and Monterossa,[26] Simpson,[27] and Garcia-Webb[28] considered the relationship between depression and hypertension and found that the treatment of one may induce the occurrence of the other. In several prospective studies depression has been found useful for prediction. Nadelson et al. found that three different measures of depression could differentially predict functional as opposed to organic disorders with a 90 to 95 per cent correct classification rate.[29]

Design

The three interventions were given in a sequential factorial experimental design.

This allows each intervention or any combination of the three interventions to be evaluated. The first intervention (treatment 1, Fig. 1) concerned clarification of the prescribed regimen, the second (treatment 2) attached social support by intervening with a family member, and the third (treatment 3) attempted to increase self confidence in compliance by changing the locus of control from an external orientation to an internal orientation. The sequential combination of interventions allowed each treatment or combination of treatments to be evaluated against a control group.

In a 2-month period, 402 patients that were identified as hypertensive from the General Medical and Hypertension Clinics of the Johns Hopkins Hospital were selected for the study. All the patients were outpatients that had been in treatment for at least 6 months. During the randomization process the experimental group was given the first intervention. The treatment 1 session was a reteaching and reinforcement of the medical practitioner encounter that immediately followed the medical intervention. All patients were identified in the waiting room and asked if they could stay 15 to 20 minutes after seeing the doctor and/or nurse to discuss their treatment for hypertension. If the patient was assigned to the control group, he or she was informed that no one was available at this time, but that they may be contacted in the future. Of the 407 patients eligible for the study, only five refused to participate, 202 received treatment 1.

Two months after the first intervention was initiated half of the control and experimental patients (200) were selected for interviewing in their homes. The average interview lasted 53 minutes. At that time, the name of a family member and permission to contact that person was obtained. The family member was selected by the patient based upon the question, "Who is the adult person in your household whom you talk with the most?" If the patient identified more than one person then he or she was asked with whom they spend the most time.

The second intervention (treatment 2) was an outreach effort in the patient's home. The family members were visited 3 months after the initial interview. They were given specific information on hypertension, what the patient should do, and special advice on what they could do to help. The sessions lasted an average of 30 minutes. Fass et al.[6] has described the intervention used in this study.

The third intervention (treatment 3) consisted of three 1.5 hour group awareness sessions. The purpose of this intervention was to increase the patients' self-confidence about controlling their own health and to change an individual from an external locus of control to an internal locus of control. Individual support was provided through the group discussion. Random sets of 16 patients were recruited for each set of group sessions. The sessions contained role playing, behavioral rehearsal, problem clarification, and general discussion of management and compliance in areas where they lacked self-confidence about their problems with hypertension. Bowler has discussed the rationale for the intervention and content of each session in detail.[30] Of the 200 patients selected, only 24 refused to participate. The sessions at Johns Hopkins Hospital were arranged to meet the patient's convenience and free transportation was provided.

The average age of the sample population was 53.7 years, 75 per cent female, and 93 per cent black. The average monthly income was about $330.00. The average education was between eighth and ninth grade. Twenty-three per cent of the population lived alone and fifty-five per cent had the cost of their medication covered by medical assistance. The effectiveness of the randomization of patients into unbiased groups was tested according to differences in sex, race, age, educational level, occupation, income, medication compliance (Table 1), and medication being paid for by medical assistance. No significant differences were found between the groups for selected baseline socioeconomic and medical variables after performing a chi-square test.

Methods

The decision process consisted of two steps. First, the variables exhibiting strong discriminating properties associated with participation in a health education intervention were selected mathematically. The results are presented in Tables 2–4. Discriminant analysis was performed to separate patients into high and low compliers within treatment groups. These results are available from the authors, but are not presented here because they essentially are in agreement with the compliance likelihood ratio tests presented in step 2 of the model. Depression, as measured by a self-report scale, was the first variable to enter the discriminant analysis, indicating that it contributed the most to separation of the groups.

In the second step, the likelihood of success (medication adherence) for experimental versus control groups (no health education effort) was developed for each of these selected discriminator variables and tested for their statistical significance. The 29 variables under consideration in step 2 are listed in Table 5. The proportion of compliers with a certain attribute was compared for the experimental versus control group. This ratio of proportions was formed and tested statistically. This ratio gives a direct estimate of the likelihood of success for one treatment versus another treatment. A likelihood ratio greater than 1 indicates that the health education intervention was effective in stimulating increased adherence to treatment. This ap-

Table 2. Independent Variables that Delineate Low and High Compliance in the Discriminant Function Analysis

Variables	Function	Related to
Age	−0.22	High compliance
Live with someone	0.25	Low compliance
Speak to someone for advice before the doctor	0.24	Low compliance
External locus of control	−0.43	High compliance
Luck is unimportant	0.21	Low compliance
Stay in hospital for HBP	0.14	Low compliance
Health is poor (perception)	0.17	Low compliance
Not worried about health	0.25	Low compliance
Control over life	0.17	Low compliance
High patient perception of staff's expectations for appointment keeping	−0.23	High compliance
Long clinic waiting time	0.40	Low compliance
Understood doctor/nurse	0.28	Low compliance
Believe that medicine works for HBP	−0.24	High compliance
Belief that HBP is serious	0.25	Low compliance
Belief in seriousness of your condition	−0.36	High compliance
Phase II—control C2. Figure 1)	0.22	Low compliance

Low compliance: Centroids, 0.43; N = 106. High Compliance: Centroids, −0.59; N = 77.
Recognition Classification Rate: 75% (for all these stable variables).
Canonical Correlation: 0.51.

TABLE 3. Variables That Indicate Participation in Family Reinforcement (Treatment 2)

Variables	Function C	Related to
Phase I—Control	−0.69	Nonparticipation
Dissatisfaction	+0.58	Participation
*Someone for advice before doctor	+0.27	Participation
Dreams about HBP complications	−0.51	Nonparticipation
Understood doctor/nurse	+0.24	Participation
Depression	+0.45	Participation
Influence over things	+0.22	Participation
*Control over life	−0.18	Nonparticipation
Family does not help follow doctor's advice	−0.28	Nonparticipation
*Health is poor (perception)	−0.26	Nonparticipation
Sex—female	+0.22	Participation
*Race—white	Positive	Participation
*Medical assistance does not cover medication cost	Positive	Participation
*Education	Negative	Nonparticipation

Participation: Centroids, +0.687; N = 35. Nonparticipation: Centroids, −0.617, N = 39.
Recognition Classification Rate: 81% (for all variables).
Recognition Classification Rate: 68% (for stable variables only).
Canonical Correlation: 0.66.
* Discriminating variables when only stable variables are considered.

proach is analogous to the criteria used in sequential decision making (Flagle[31]).

Linear discriminant functions are effective for both prediction and description of decision processes. Discriminant functions recently have been applied to predict medical problems. Thomas and Greenstreet used discriminant analysis in studying psychobiological characteristics in youth as predictors of five disease states: suicide, mental illness, hypertension, heart disease, and cancer.[32] After the success of decision modeling in the area of diagnosis classification,[33] it would seem logical to extend decision modeling into treatment assignment.

The discriminant analysis has several steps. First, variables are selected for inclusion in the analysis based upon a criterion that deals with group separation, in our study, the criterion maximizing the square root of the distance between the two closest groups.[34] Before a variable is selected it has to meet another criterion, that of increasing total group separation. The first variable selected has the highest value on the criterion. Then the single variable is paired with each of the other available variables, one at a time, and the selected criterion computed. The new variable, in conjunction with the old variable, which has the highest value on the criterion, is selected. The process is continued until no improvement is made or a minimal level is reached, e.g., the variable is considered for selection if its partial multivariate F ratio is larger than a specific value. At the beginning of each variable selection step, each previously selected variable is checked to see if it still contributes to the discrimination. If a variable loses its discriminating power, it is removed, e.g., the multivariate F ratio is less than a certain value.

TABLE 4. Variables That Indicate Participation in Small Group Meeting (Treatment 3)

Variables	Function C	Related to
*Age	+0.56	Participation
Family does not help follow doctor's advice	+0.43	Participation
*Dreams about HBP complications	+0.41	Participation
Belief in benefit	+0.41	Participation
Phase I—Control (C1, Figure 1)	−0.32	Nonparticipation
Belief in seriousness	+0.52	Nonparticipation
Understood doctor/nurse	+0.35	Participation
*Patient worried about HBP	−0.24	Nonparticipation
Race—white	−0.35	Nonparticipation
Medicines work for HBP	+0.39	Participation
Dissatisfaction	+0.34	Participation
High blood pressure is serious	+0.37	Participation
Influence over things	+0.20	Participation
*Education	+0.26	Participation
Live with someone	+0.17	Participation
*Patient's perception of staff's expectations of patient's appointment keeping (high)	Negative	Nonparticipation

Participation: Centroids, +0.531; N = 58. Nonparticipation: Centroids, −0.670; N = 46.
Recognition Classification Rate: 75% (for all variables).
Recognition Classification Rate: 67% (for stable variables only).
Canonical Correlation: 0.60.
* Discriminating variables when only stable variables are considered.

After the important variables are selected, the corresponding matrix is solved to find a weighted function of the variables such that the F ratio for the weighted (w) mean scores (x) for the samples in a one-way analysis of variance is maximized. The problem can be formulated in the matrix algebra as the set of weights that maximizes the expression:

$$\lambda = (w'\ W\ s)/(w'\ B\ w),$$

where B = between groups' p × p matrix of sums of squares and cross-products of the x scores about their respective means,
W = the corresponding within group matrix,

and the p × 1 column of weights is represented by vector w (with transpose w'). Since the expression involves scalar quantities and, if the total size is N, the degrees of freedom for "between" and "within" groups are (K-1) and (N-K), respectively,

$$\lambda = \frac{K-1}{N-K}\,F.$$

Therefore, maximizing λ is equivalent to maximizing the F-ratio between groups and thus maximizing the difference between group means.

The first function in the process will explain the largest share of explained variance, by definition. The relative contribution of each variable to the respective function can be obtained from the weights for the standardized discriminant function. In the standardized function, over all cases, the scores for one function will have a mean of 0 and a standard deviation of 1. If a weight has a value of 0.70 then the variable is twice as important as a variable with a 0.35 value. Nonstandardized coefficients do not represent relative importance because they have not been adjusted for the measurement scales and variability in the original variables.‡

Scales previously have been developed for measurement of the adherence to medication dependent variable and the six independent variables: depression, dissatisfaction, belief in seriousness, belief in benefits, health external locus of control, and patient's perception of staff's expectations for patient's appointment keeping. The reliability coefficient for measuring the internal consistency of the times used in the scale is Cronbach Alpha.[35] The previously published[7] adherence to medication scale had a Cronbach Alpha of 0.92 with four items, i.e., there is a high tendency toward consistency from one question to the next. The depression scale had a high reliability of 0.85 with seven items.

A complex construct, such as depression, reflects a broad range of signs and symptoms of the clinical depressive syndrome. Symptoms of dysphoric affect and mood, withdrawal of interest in life activities, lack of motivation, and loss of vital energy are all depressive symptoms. In addition, feelings of helplessness and futility also are key signs. The first six items in the depression scale come from the SCL-90 Hopkins Symptom Checklist scale reported by Derogatis.[36] The patient was asked to read a list of problems and complaints and "report how much discomfort that problem has caused during the past year" (feeling blue, lonely, hopeless, worthless, suicidal, or having no interest in things). A seventh item was included in the depression scale and targeted specifically to the hypertensive patients: "Do you have feelings of hopelessness about your high blood pressure?"§

Results

The compliance scale consists of four adherence to medication items: ever forget your medicine, careless at times about taking your medicine, sometimes stop taking your medicine if you feel better, and sometimes stop taking your medicine if you start to feel worse. Table 2 summarizes the results of the discriminant function analysis for high and low compliance. The results indicate that factors such as age, believing in the medication, staff expectation for keeping all appointments, and belief in the seriousness of hypertension all contributed to higher compliance behavior.

Tables 3 and 4 summarize the results of the discriminant function analysis of "participation in the family member reinforcement" and "participation in the group meeting," respectively. In Table 3, the function measures "dissatisfaction" and "depression" on the positive (higher) end, i.e., indicating higher participation in the family reinforcement therapy (intervention 2). The positive scores relate to participation in the family-member reinforcement, the negative scores to nonparticipation. The recognition classification rate is the percentage of the cases that the model correctly classified based upon preassigned classifications. The recognition classification rate of 81 per cent is considerably over the 50 per cent that would be expected by chance.

---

‡ Another measure for importance of a function is the canonic correlation, which is derived from the eigenvalues (latent roots). The canonic correlation represents the ability of each function to separate groups by the per cent of variance contained in the discriminating variables that is explained by the function. The canonic correlation squared is the proportion of variance in the discriminant function explained by the groups. The centroids for each function are the mean of each group for that function. Comparison between centroids represents the separation of the groups.

§ The questionnaires used in the development of the scales are available from the authors upon written request. The depression scale has been validated against other self-report instruments and scales, such as the Minnesota Multiphasic Personality Inventory and the Zung inventory for measuring clinical depression,[37] but has yet to be validated against any depression scale based on paraprofessional interviews (e.g., the Hamilton inventory).[38]

TABLE 5. Exogeneous and Endogeneous Independent Variables

Exogeneous Variables (Stable and not affected by the intervention)	Endogeneous Variables (Variables that are dependent to some unknown extent on the intervention)
Sex—female	Family does not help follow doctor's advice
Race—white	Dissatisfaction with the medical care system scale
Age	Belief in seriousness scale
Working status	Believe that high blood pressure is serious
Education	Believe that medicines work for high blood pressure
Medical assistance does not cover drug cost	Don't know enough to care for HBP
Live with someone	Understood doctor/nurse
See someone for advice before doctor	Treatment 1
Perception of long clinic waiting time	Treatment 2
Health external locus of control scale	Treatment 3
Control over life (ladder scale)	
Influence over things	
Luck is unimportant	
Depression scale	
Stay in hospital for high blood pressure	
Dreams about HBP complications	
Patient worried about high blood pressure	
Health is poor (perception)	
Not worried about health	
Perception of staff's expectations of patient's appointment keeping (scale)	

The exogenous and endogenous independent variables considered in the analysis of adherence to medication are given in Table 5. Tables 6 and 7 contain the decision rules of the likelihood ratio test. The results from the analyses designed to discriminate compliers from noncompliers are utilized for determining the significant variables. We then compute likelihood ratios,[‖] and consider the interrelationships of the health education interventions being evaluated. Statistically significant results were found with the likelihoods developed for the 18 selected independent variables (Table 7). The consistency between the results of the discriminant analyses and the likelihood calculations corroborate the results.

The first intervention produced an inverse effect between compliance and depression (Table 6). The first intervention may be effective for individual's with no depression. It may be ineffective for those individuals with high depression and those individuals who are not worried about their health. The first intervention is a reinforcement of the doctor's instruction, which allows patients to take better care of themselves; this is an active medical reg-

---

[‖] A likelihood can be constructed from any two probabilities or proportions. The likelihood ratio of high compliance versus low compliance for a treatment may be interesting but it cannot be statistically tested for difference in proportions. For two proportions to be tested statistically with a binomial "t" test, they must be independent. Thus, for the likelihood ratios to meet this criteria, they must compare proportions from groups that are independent. This was accomplished in our study by the use of the randomized experimental design approach.

TABLE 6. Adherence to Medication Likelihood as a Function of Depression

Treatment Phase(s)	Given: High Depression*			Given: Medium Depression			Given: Nondepression		
	Likelihood of High Compliance†	Significance		Likelihood of high Compliance	Significance		Likelihood of High Compliance	Significance	
		t	p		t	p		t	p
Phase I E1 vs. C1	0.66	−1.32	<0.10	0.91	−0.52	—	1.52	1.32	<0.10
Phase II E22 vs. C22	2.50‡	2.54	<0.01	1.13	0.52	—	0.42		
Phase III E34 vs. C34	0.36			1.52	1.09	<0.20	0.25		
Phase I & II E21 vs. C21	0.73	−0.59	—	1.76‡	2.46	<0.01	1.86‡	1.83	<0.05
Phase II & III E33 vs. C33	0.90			0.89	−0.35	—	§		
Phase I & III E32 vs. C32	1.00			1.74	1.34	<0.10	1.61		
Phase I, II & III E31 vs. C31	0.88			0.93		0.25		0.68	

* No depression score is 7, a medium depression is greater than 7 to and including 16, a high depression is greater than 16 to and including 35.

† A low compliance is 5 to and including 13 and a high compliance is greater than 13 to and including 15. There were natural break points in the response distribution.

‡ Significant at the 5 per cent level. Sample sizes of high compliers in these three cases, reading left to right, are 16, 40, and 17, respectively.

§ All cases were high compliers.

TABLE 7. Likelihood of Success in Adherence to Medications As a Function of Type of Treatment

Interventions*	Variables	Likelihood*	Percentage Increase in High Compliers	Significance†	Actual Sample Size
Treatment 2 E22 vs. C22	Depression (high)	2.46	156%	0.01	(11/17)/(5/19)
	Health is fair/poor	1.58	58%	0.05	(8/23)/(11/25)
Treatment 3 E34 vs. C34	Education (9th grade or more)	0.48	−52%	0.05	(6/20)/(8/13)
	Health is excellent/good	0.47	−53%	0.05	(6/18)/(5/17)
Treatment 1 & 2 E21 vs. C22	Medical assistance pays for medicine	1.54	54%	0.09	(24/40)/(16/41)
	Clinic waiting time (all right)	1.54	54%	0.05	(24/32)/(17/35)
	Age (19 to 53)	2.0	100%	0.01	(20/39)/(10/39)
	Not working	1.55	55%	0.05	(30/47)/(21/51)
	Not worried about health (little or not at all)	2.04	104%	0.002	(27/53)/(15/60)
	Health external LOC	1.83	83%	0.05	(16/35)/(7/23)
Treatment 2 & 3 E33 vs. C34	Education (less than 9th grade)	0.48	−52%	0.05	(5/17)/(8/13)
	Education (9th grade or more)	2.13	1.95	0.05	(10/17)/(5/18)
	Says health is good/excellent	0.43	−57%	0.05	(5/16)/(5/7)
	Says health is fair/poor	1.79	79%	0.05	(11/20)/(8/26)
Treatment 1 & 3 E32 vs. C34	Education (less than 9th grade)	0.39	−71%	0.05	(5/21)/(8/13)
	Education (9th grade or more)	1.97	97%	0.05	(12/22)/(5/18)
	Says health is excellent/good	0.43	−58%	0.05	(5/18)/(9/14)
Treatment 1, 2 & 3 E31 vs. C34	Clinic waiting time (all right)	1.70	70%	0.05	(13/17)/(8/15)

* Refer to Figure 1

† All but one item in this column are significant at the 5 per cent level or better.

imen. One possible explanation is the state of "no depression" is related to being receptive to active medical regimens and the reverse is true for high levels of depression.

The second intervention was effective for patients with high self-reported levels of depression and the feeling that one's health is fair or poor (Table 7). Depression was not important in the discriminant analysis without the inclusion of the other variables. One rationale for explaining why depression is an important characteristic for indicating responsiveness to the second intervention is that a depressed person generally tends to be passive and should be influenced by family members who have learned how to reinforce the patient in the intervention.

The decision rules listed in Table 7 indicate that a combination of the first and second interventions are best, in general. For the practitioner this would mean that everyone should receive clarification of the doctor's message and family member reinforcement that can be tailored to the patient and his family member (treatment phase 1 and 2). There are exceptions to the superiority of treatments 1 and 2. The highly depressed patient does better with only reinforcement of the family member (treatment 2), which is consistent with the condition of depression and related passivity. Patients with more than a 9th grade education do better with intervention 3 combined with one or two. Surprisingly, combinations of interventions had a detrimental (decreased adherence) effect on patients with less than a 9th grade education.

The most important limitation in generalizing these results is the sample population. As the results are extended beyond this temporal point and population, the associated limitations must be considered. In a more technical vein, multicollinearity between patient characteristics critical for triaging patients is a limited problem considering the low Pearson correlation coefficients. Depression was seldom significantly related to "health is poor (perception)," "not worried about health," "control over life," and "age" (0.27, −0.38, −0.18, and −0.18 correlation coefficients, respectively). "Age" was significantly related to "education" and "live with someone" (correlation coefficients −0.46 and −0.24, respectively). Older people tend to live alone and have less education. There were four variables that were considered important in the design phase but were not analyzed because the response rate by patients was too low. These were: "occupational level," "income," "medicine cost is OK," and "visit cost is OK."

One final caveat is in order. If the sample size had been larger, a holdout sample could have been used to test the predictive ability of the descriptive analyses presented. Such a technique should be employed with data from the sample of 3,400 Australian hypertensives that were randomized into either therapy group or placebo.[39]

Weighing Costs and Benefits

A limited benefit–cost comparison was developed to evaluate the results of the triage methods. Benefits had to be defined more broadly than simply the employer benefits of hypertension health education projects accrued with a 5-year time period.[40, 41] The model compared the costs of the intervention(s) versus the total savings from the intervention(s). Two types of benefit, prevented (foregone) indirect earnings loss and direct medical costs, were considered for our study. The first benefit measure was the estimated preserved earnings per person having a diastolic blood pressure less than 94 mm Hg versus 95 mm Hg or greater. The second benefit measure was the savings in expected annual cost per person for treatment of cardiovascular disease and nephritis (excluding cost of hypertension treatment per se) for a diastolic blood pressure below 94 mm Hg versus a level above 95 mm Hg. The benefit-adherence ratio coefficient came from a 1975 study by Stokes and Carmichael.[42] The data was inflated to 1979 dollars by the medical price index component of the consumer price index and earning trend index.[43]

The cost figures were adjusted for age and sex in the study population with the exception that the figures for the population under 35 years of age (9.5 per cent) were applied to the 35–44 age data and the figures for population over 64 years of age (25.6 per cent) were applied to the 55–64 age data because the 1975 raw data were not available. This approach produced a conservative estimate of savings since the population over 64 years of age would have a higher associated savings. Problems in underestimating the benefits and costs of personal medical care[44] and health programs[45] have been reviewed in detail elsewhere. The other assumption was that being defined as a high complier (42.4 per cent of the study population) or a low complier (57.8 per cent of the study population) is perfectly correlated with blood pressure control. In analyzing the cost of sequelae, blood pressure control was defined as a diastolic between 85 mm Hg and 94 mm Hg, and not being in control was defined as a diastolic blood pressure between 95 mm Hg and 104 mm Hg.

The costs of the treatments were estimated by adding the marginal personnel cost to provide each intervention. The cost of each intervention annually was estimated to be $5.35 for the clarification of the doctor's message; $19.14 for the family member reinforcement; and $17.00 for the series of three group meetings. The annual savings in earnings loss per person for control versus noncontrol was $53.00. The savings in expected annual cost per person for treatment of cardiovascular disease and nephritis for control versus noncontrol was $56.00. The limited benefit–cost model for the health education triage methods including the cost of the health education interventions is:

$$R = \frac{IHC \times (S_1 + S_2)}{PC} =$$

Benefit to cost ratio for intervention programs.

IHC = increase in number of additional high compliers with treatment relative to the control group (per 100 patients).

S_1 = savings per high complier in earnings loss per person for coantrol group versus noncontrol group.

S_2 = savings per high complier in expected annual cost for treatment of cardiovascular disease and nephritis for control group versus noncontrol group.

PC = additional treatment cost for intervention program versus control group.

Given the assumptions, the cost and savings information, and the effectiveness of the intervention programs, the decision to triage is most cost-beneficial for the 24 per cent highly depressed population (defined in Table 6). The benefit–cost ratio of 1.24 for hypertension control for persons in the age range of 35–65 (Option 4, Table 8). In other words, selecting the 24 per cent highly depressed subpopulation and providing family member reinforcement to this group is more cost beneficial than giving everyone the special health education intervention. Previous studies demonstrating a cost–benefit ratio in the 1.1–1.3 range may not stand the test of time in claiming statistic significance above the 1.0 level when applied to a larger population, i.e., less susceptible to the Hawthorne effect.[46] In other words, the newness and small scale of the program and the extent to which the patients are aware that they are part of a research project, may make the degree of program effectiveness less

TABLE 8. Simple Benefit-To-Cost Comparisons of Triage Options Versus the Option Not to Triage

Option	Triage	Type of Patients Triaged	IHC*	Health Education Intervention(s)	Benefit Cost Ratio
1	Yes	High level of depression	65–26	Family reinforcement (FR)	2.20†
2	Yes	Medium level of depression	58–33	FR + message clarification	1.15
3	Yes	No depression	65–35	FR + message clarification	1.33
4	No	All patients	60–32	FR + message clarification	1.24

* IHC = increase in number of high compliers with treatment versus control per hundred patients.
† Only option 1 has a significantly better ratio for triaging in comparison to not triaging.

generalizable to future situations. These results must be considered preliminary because of the lack of long-term follow-up. A 5-year following and larger sample size also would strengthen confidence in predictive applications of the discriminate analysis.

Conclusions

The premise of this study is that patients will derive a greater benefit from a health care system that assigns them to a health education treatment that is more effective for them. This assumes that there are certain characteristics of the patient that predispose him to respond better to one treatment than another and that these characteristics are measureable.[47] The decision rule in our model (Table 7) is to assign patients into health education treatments based on likelihoods of success statistically significant at the 5 per cent level. For example, the characteristic depression, as measured by the response to seven questions, is found to be significant in explaining treatment compliance. The effectiveness of the family reinforcement intervention for increasing the adherence to medication is statistically effective for highly depressed patients. Considering this type of differential effectiveness for patients with various characteristics, the ability to assign patients into health education intervention has the potential to maximize the achievable benefits of hypertension detection efforts. Future research should consider whether the annual cost of these special interventions continues to yield the same annual benefits, diminishing benefits, or increased benefits.

More precise data linking adherence to symptom decline and to improved health would provide greater accuracy for policy makers. The actual relationship between our simple high and low dichotomy for compliance and diastolic blood pressure levels also should be established definitively. However, no matter how refined the model, many dynamics of the situation could not be considered given our present poor state of information concerning outcomes of hypertension treatment. Better epidemiologic studies linking symptom reduction (blood pressure decline) to better health are necessary.[48] The major finding of our analysis is the potential importance of one mental health variable in explaining the effectiveness of health education efforts. The results in the depression scale would encourage the inclusion of other standardized measures of mental health symptoms in future studies.

References

1. Green L. Determining the impact and effectiveness of health education as it relates to federal policy. Health Educ Monogr 1978;6:28.

2. Eastaugh S. Cost of elective surgery and utilization of ancillary services in teaching hospital. Health Serv Res 1979;14:290.

3. Green LW, Levine DM, Deeds S. Clinical trials of health education for hypertensive outpatients: design and baseline data. Prev Med 1975;4:417.

4. Glanz K, Kirscht J, Rosenstock I. Linking research and practive in patient education for hypertension. Med Care 1981;19:141.

5. Conte A, Brandzel M, Whitehead S. Group work with hypertensive patients. Am J Nurs 1974;5:910.

6. Fass, MF, Green LW, Levine DM. Design of an experimental intervention to increase family reinforcement in the home for patient compliance. Presented at the American Public Health Association, Chicago, November 1975.

7. Levine D, Green L, Deeds S, Chawlalo J, Russell R, Finlay J. Health education for hypertensive patients. JAMA 1979;241:1700.

8. Morris L, Halpern J. Effects of written drug information on patient knowledge and compliance: a literature review. Am J Public Health 1979;69:47.

9. Dickinson J, Warshaw G, Gehlbach S, Bobula J, Muhlbaier L, Parkerson G. Improving hypertension control: impact of computer feedback and physician education. Med Care 1981;19:843.

10. Eastaugh S. Teaching the principles of cost-effective clinical decision-making to medical students. Inquiry 1981;18:28.

11. Mitchell JH. Compliance with medical regimens: an annotated bibliography. Health Educ Monogr 1974;2:75.

12. Green L. Toward cost-benefit evaluations of health education: some concepts, methods, and examples. Health Educ Monogr 1974;2:34.

13. Kasl SV. The health education belief model and behavior related to chronic illness. Health Educ Monogr 1974;2:433.

14. Battistella R. Disease prevention and health education. In: Battistella R, Rundall T, eds. Health care policy in a changing environment. Berkeley: McCutchan, 1978:274.

15. U.S. Department of Health, Education and Welfare. Health—United States 1976–1977. HEW Publication (HRA) 77-1232, Government Printing Office, Washington, D.C., 1977.

16. Breslow L. Risk factor intervention for health maintenance. Science 1978;200:908.

17. Lew EE. High blood pressure, other risk factors and longevity: the insurance viewpoint. Am J Med 1973;55:281.

18. Weinstein M, Stason W. Hypertension: a policy perspective. Cambridge: Harvard University Press, 1976.

19. Wagner EH, Warner JT, Slome C. Medical care use and hypertension." Med Care 1980;18:1241.

20. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertensive II: results in patients with diastolic blood pressure averaging 90 through 114 mm Hg. JAMA 1970;213:1143.

21. Freis E. The clinical spectrum of essential hypertension. Arch Intern Med 1974;133:982.

22. Langfeld SBl. Hypertension: deficient care of the medically served. Ann Intern Med 1973;78:19.

23. Junod B, Gutzwiller F. Community based cardiovascular risk factor experiment. Medecine Sociale et Preventitive 1977;22:157.

24. Nelson EC, Stason WB, Neutra RR, Soloman HS, McArdle PJ. Impact of patient perceptions on compliance with treatment for hypertension. Med Care 1978;16:893.

25. Naditch MP. Locus of control, relative discontent and hypertension. Social Psychiatry 1974;9:111.

25a. Lau R, Ware J. Refinements in the measurement of health-specific locus of control beliefs. Med Care 1981;19:1147.

26. Bloom J, Monterossa S. Hypertension labeling and sense of well-being. Am J Public Health 1981;71:1228.

27. Simpson FO. Hypertension and depression and their treatment. Aust NZ J Psychiatry 1973;7:133.

28. Garcia-Webb P. Hypertension and depression. Aust NZJ Psychiatry 1973;7:298.

29. Nadelson T, Flatley M, Lipsitt D. Quantitative indicators of functional disease. Massachusetts Institute of Technology Technical Report ACP-19, Lincoln Laboratory Project Report Series, Lexington, Massachusetts, 1971.

30. Bowler M. Internality training sessions with hypertensive patients. Working paper, Baltimore, Md., Johns Hopkins Hospital, 1977.

31. Flagle CD. A decision theoretical comparison of three procedures of screening for a single disease. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol IV, Biology and Problems of Health, 1966. Lecam L, Neyman J, eds. Berkeley: University of California, 1966.

32. Thomas CB, Greenstreet RL. Psychobiological characteristics in youth as predictors of five disease states. Johns Hopkins Med J 1973;132:16.

33. Taylor TR, Shields S, Black R. Study of cost-conscious computer-assisted diagnosis in thyroid disease. Lancet 1972,2:79.

34. Maxwell AE. Multivariate analysis. In: Holland W. ed. Data handling in epidemiology. London: Oxford, 1970:149.

35. Stanley JC. Reliability. In: Thorndike RL, ed. Education measurement. Washington, DC: American Council on Education, 1971:356.

36. Derogatis L, Lipman R, Rickles K, Uhlenhuth E, Cori L. The Hopkins symptom checklist: a self-report symptom inventory. Behav Sci 1974;19:1.

37. Derogatis L, Rickels K, Rock A. The SCL-90 and the MMPI: validation of a new self-report scale. Br J Psychiatry 1979;128:280.

38. Freedman A, Kaplan H, Sadock B, ed. Modern synopsis of the comprehensive textbook of psychiatry—II. Baltimore: Williams and Wilkens, 1976:1012.

39. Australian Therapeutic Trial in Mild Hypertension. Report by the ATTMH management committee. Lancet 1980;1:1261.

40. Hannan E, Graham J. A cost-benefit study of a hypertension screening and treatment program at the work setting. Inquiry 1978;15:345.

41. Haurer J, Goodman J. Evaluation of performance and cost in a hypertension control program. Med Care 1980;18:485.

42. Stokes J, Carmichael DC. Cost benefit analysis of model hypertension control. National High Blood Pressure Education Program, National Heart and Lung Institute, National Institute of Health, Bethesda, Md., 1975:32.

43. United States Bureau of Labor Statistics. Monthly Labor Review 1979;102(12):71.

44. Eastaugh S. Medical economics and health finance 1981. Boston: Auburn, pp. 55.

45. Eastaugh S. Effectiveness of community-based hospital planning: some recent evidence. Applied Economics 1982;14:(in press).

46. Shortell S, Richardson W. Health program evaluation. St. Louis: CV Mosby, 1978:47.

47. Williamson JW. Assessing and improving health care outcomes: the health accounting approach to quality assurance. Cambridge, Mass.: Ballinger, 1978.

48. Szklo M. The epidemiologic basis for hypertension control. Journal of Public Health Policy 1980;1:312.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Social Networks, Beliefs, and the Decision to Seek Medical Care: An Analysis of Congruent and Incongruent Patterns

EMIL BERKANOVIC, PH.D., AND CAROL TELESKY, M.A.

This article focuses on the impact of congruent and incongruent patterns of symptom specific beliefs and social network advice on the decision to seek medical care for symptoms. Data from 769 individuals who reported symptoms in the Los Angeles Health Survey are used to examine this issue. Further, the role of more general health orientations and social network influences in the decision to seek care is examined under these congruent and incongruent conditions. It is concluded that the more general factors, which have little effect on the aggregate, may play a role in decision making when beliefs and advice are incongruent. Further, it is concluded that reports of beliefs and advice about specific symptoms that have been given after the decision has been made to seek care may be valid and not the result of retrospective reconstruction.

IN A RECENT ARTICLE in Medical Care,[1] we examined the degree to which need, social structural, organizational, social network, and social psychological variables could predict the decision to seek medical care for symptoms. In that analysis, the independent variables were categorized into five main groups:

1. Need factors, including the characteristics of the symptom under study and the individual's overall health status.
2. Social structural factors that might predispose or enable individuals to seek medical care when confronted by symptoms.
3. Organizational factors that might enable individuals to seek medical care.
4. Social network patterns and general health orientations that might predispose individuals to seek care.
5. Specific social network influences and personal beliefs about the symptoms under study.

In a hierarchic multiple regression analysis, this set of independent variables explained 57 per cent of the variance in the decision to seek medical care for the symptoms reported over a study year. Of the 57 per cent of the variance explained, however, 41 per cent was explained by symptom-specific network advice and beliefs. Further, of the remaining 16 per cent, 12 per cent was explained by the need variables. These data are summarized in Table 1.

Because of the distribution of variance we concluded that the analysis reported in Table 1 was difficult to interpret. Specifically, the failure of the structural variables, the social network patterns, and the

---

From the School of Public Health, University of California, Los Angeles, California.

Supported by Grant 5-R18-CA18451, awarded by the National Cancer Institute, Department of Health and Human Services. A preliminary version of this article was presented to the International Society of Applied Anthropology, April 15, 1981, Edinburgh.

0025-7079/82/1000/1018/$01.25 © J. B. Lippincott Co.

TABLE 1. Hierarchial Multiple Regression Analysis of the Decision to Seek Medical Care for Symptoms

Step	Variables	Standardized Regression Coefficients*
I. Indicators of need	Subjective health status	0.05
	Number of chronic health problems	0.12†
	Reported disability from symptom	0.28‡
	$r^2 = 0.116$	
II. Sociodemographics	Total family income	0.05
	Socioeconomic index	−0.03
	Race/ethnicity (white vs. other)	−0.01
	Marital status (now married vs. other)	−0.02
	Sex (female = 1, male = 0)	0.04
	Age	0.09§
	$r^2 = 0.125$	
III. Organization of care	Continuity of care (index)	0.08§
	$r^2 = 0.130$	
IV. Social network characteristics and health orientations	Amount of network contact	−0.10§
	Amount of network consulting	−0.08
	Network support	0.11§
	Network size—total number of members	0.00
	Network advice to see doctor	−0.01
	Acceptability	0.02
	Accessibility	0.02
	Seriousness of illness	−0.02
	Perceived efficacy of care	−0.01
	Susceptibility to illness	0.06
	Motivation—concern with health	0.05
	Concern with medical costs	−0.03
	Tendency to delay seeking care	−0.05
	$r^2 = 0.151$	
V. Symptom specific networks and beliefs	Amount of network contact (log)	0.02
	Distance of network members consulted	0.11†
	Perceived influence of network	0.03
	Size of network consulted	0.02
	Network advice to see doctor	0.15†
	Perceived likelihood of recurrence	−0.06§
	Perceived efficacy of care	0.51‡
	Ease of going to doctor	0.09‡
	Perceived seriousness of symptom (index)	0.17‡
	$r^2 = 0.568$	

* Regression coefficients shown in Table 5 are presented in hierarchic fashion. Each coefficient represents the amount of change in the dependent variable that is attributed to a 1 unit change in the standard score of the predictor variable, independently of all other variables examined in the same or prior steps.

† $p < 0.01$.

‡ $p < 0.001$.

§ $p < 0.05$.

general health orientations to account for much of the variance in the analysis raised two troublesome questions. First, these findings failed to support the notion that behavior is linked to both social structure and enduring cognitive orientations. Thus, although substantial variance is explained, these data provide little insight into how social structure and prior experience, as expressed through cognitive orientations, impact on responses to individual symptoms. Second, since the symptom-specific data were collected subsequent to the decision to seek medical care, the validity of these findings may be jeopardized because of retrospective reconstruction. The issue of retrospective reconstruction is especially important because it is unlikely that symptom responses can be studied prospectively in the general population. To do so would require collecting data from the moment the individual becomes aware of a symptom.

The purpose of this article is to explore these effects further. Specifically, this analysis focuses on the impact of congruent and incongruent patterns of symptom specific beliefs and social network advice in the decision to seek medical care for these symptoms. In addition to providing further insight into how these symptom-specific factors combine to affect the decision to seek care, this analysis provides an opportunity for examining how variables that appear to have little effect on the analysis reported in Table 1 impact this decision under one or another of the combination of beliefs and network advice. Further, by comparing the impact of congruent and incongruent patterns, it may be possible to make some inference regarding the extent to which retrospective reconstruction is a viable interpretation of the results in Table 1.

Sample

The data to be reported were collected during a 1-year panel study as part of the 1976–1977 Los Angeles Health Survey (LAHS). The sample design for the LAHS was a three-stage random probability sample of Los Angeles County that was developed by the Institute for Social Science Research (ISSR) at UCLA.[2] From October through December 1976, demographic and health data were obtained from 1,210 respondents during the initial face-to-face interviews. The initial interview then was followed by seven telephone reinterviews at approximately 6–8 week intervals and a final face-to-face interview at the end of the panel year. Thus, 1 year of symptom experience is being analyzed in relation to the background information collected during the first interview. Details of variable measurement and construction are reported in Berkanovic, Telesky, and Reeder[1] and do not need to be reviewed here.

For the 769 respondents who reported any symptoms during the 1-year study, a total 1,679 "most important" symptom episodes were reported, with individual respondents reporting from 1–17 symptoms. Multiple symptoms reported by the same person are not independent, which violates a key assumption of multiple regression analysis. Thus, while the symptom data base contains 1,679 observations, the sample size for significance testing should be 769. Accordingly, we use the symptoms as the unit of analysis, weighting the data inversely by the number of symptoms reported by the respondent. This results in conservative significance tests that compensate for the lack of independence of the symptoms. One of the unique features of this study is that the predictor variables can be linked with specific symptom responses, including use of health services. Therefore, when examining a particular symptom episode we know how the person perceived the symptom and which social network members were consulted regarding that symptom. This link between specific health beliefs, social network characteristics, and symptom responses is a central component of the analysis.

Two questions regarding what the person believed to be true about each symptom reported during the study year were asked: How serious did you think this condition was? When you realized you had this problem, how much good did you think a doctor could have done you? Individuals who said they had spoken to a friend or relative about a symptom were asked "What kind of advice did (. . .) give you about this health problem?" Answers ranged from "wait and see" to "get medical advice immediately." Not one respondent reported being advised against seeing a doctor. For purposes of the present analysis, the respondents reporting symptoms are divided into those who were advised by at least one person to see a physician and those who were not so advised. Thus, congruence occurs when the person's beliefs support seeking care and the person has been advised to do so, or when the person's beliefs do not support seeking medical care and the person has not been advised to do so. The other two possible combinations of these variables indicate incongruence between what the person believes and what he has been advised to do by his social network.

Results

Table 2 presents the percentage of individuals who saw a doctor for symptoms for each of the congruent and incongruent conditions: When the network advised seeing the doctor and the person believed that a doctor could help relieve the symptom, 90 per cent of those with symptoms actually went to see the physician. In contrast, when there was no network advice to see a doctor and this belief was negative only 8 per cent went to see the doctor. In the incongruent conditions, about a third of those who did not believe a doctor could do them some good but who were advised to see a doctor by their network went to the doctor. However, 69 per cent of those who believed the doctor could help relieve their symptoms but who were not advised to see a doctor went. As one would expect, it appears that behavior is largely consistent with congruent personal beliefs and network advice, that without the reinforcement of advice substantially fewer people act according to their beliefs and that network advice can lead to behavior believed to be ineffective.

TABLE 2. Per Cent of Respondents Initiating a First Doctor Visit for a Symptom by Perceived Efficacy of Care and Network Advice to Seek Care (n = number of respondents in each category)

Proportion of Network Members Advising Respondent to See Doctor for Symptom	Perceived Efficacy of Care for Symptom	
	High Perceived Efficacy	Low Perceived Efficacy
Some or all	90.5% (339)	33.5% (79)
None	68.6% (553)	8.5% (411)

Table 3 presents four, four-step hierarchic multiple regressions of the decision to seek medical care for symptoms classified according to each of the four congruent and incongruent conditions in Table 2. Looking first at the total amount of variance explained, it may be seen that while 13 per cent of the variance in the decision to seek medical care is accounted for in the two congruent conditions, and 15 per cent is accounted for when the belief exists that the doctor can help relieve the symptom but the network does not advise a visit, 39 per cent of the variance in the decision to seek medical care is accounted for when the network advises seeing a physician but the person does not believe a doctor can do much to help relieve the symptom. While roughly equivalent amounts of variance in each of these analyses are accounted for in the first three steps, the fourth step, which enters general network characteristics and general health orientations, contributes substantially more to the variance explained

TABLE 3. Standardized Regression Coefficients for the Decision to Seek Medical Care for Symptoms by Network Advice and Perceived Efficacy of Care

Step	Received Advice/ High Efficacy (n = 339)	No Advice/ Low Efficacy (n = 411)	No Advice/ High Efficacy (n = 553)	Received Advice/ Low Efficacy (n = 79)
I. Indicators of need				
Number of chronic conditions	0.00	0.20*	0.16*	0.05
Disability from symptom	0.14†	0.09‡	0.21*	0.25†
Subjective health status	−0.09	0.05	0.01	−0.06
r^2 for Step I	0.02	0.06	0.08	0.05
II. Sociodemographics				
Total family income	−0.12	0.02	0.04	0.26†
Socioeconomic index	0.06	−0.07	−0.02	−0.05
Race/ethnicity (white vs. other)	0.14†	0.04	0.05	0.05
Marital status (now married vs. other)	0.03	−0.04	−0.11†	−0.05
Sex (female = 1, male = 0)	−0.08	0.03	0.02	0.08
Age	0.06	0.15*	0.09	0.08
r^2 for Step II	0.06	0.09	0.10	0.12
III. Organization of care				
Continuity of care	0.10	0.02	−0.04	−0.01
r^2 for Step III	0.07	0.09	0.10	0.12
IV. Social Network Characteristics and Health Orientations				
Amount of network contact	−0.09	−0.12	−0.13†	0.03
Amount of network consulting	−0.06	−0.07	−0.10	−0.06
Network support	0.05	0.14†	0.13†	0.01
Network size—total number of members	0.02	−0.02	0.02	−0.31†
Network advice to see doctor	−0.02	−0.11†	0.02	−0.07
Acceptability	0.08	0.03	0.00	−0.08
Accessibility	0.14†	−0.01	0.02	0.01
Seriousness of illness	0.04	−0.05	0.12†	0.16
Perceived efficacy of care	−0.11	0.03	0.04	−0.12
Susceptibility to illness	0.01	0.07	0.00	0.24‡
Motivation—concern with health	0.10	0.02	0.05	−0.03
Concern with medical costs	−0.06	−0.07	−0.12†	0.25‡
Tendency to delay seeking care	−0.03	−0.02	−0.06	0.34†
Total r^2 =	0.13	0.13	0.15	0.39
n	(339)	(411)	(553)	(79)

* $p < 0.01$.
† $p < 0.05$.
‡ $p < 0.10$.

among those who do not think a doctor can do much to help relieve their symptoms, but who are advised to go by their network. Further, the factors that are statistically significant at this step are network size, in which smaller networks seem to have more influence, perceived susceptibility to illness in general, concern about the cost of care, and self-reported tendency to delay seeking treatment. The only other variables to achieve statistical significance in this analysis are disability due to the symptom and income. Thus, among those who do not believe a doctor can do much to relieve their symptoms, but who are advised by their social networks to see a physician, characteristics such as greater disability, higher income, greater perception of one's susceptibility, a smaller social network, greater concern about the costs of care, and greater recognition that one tends to delay seeking medical attention for all symptoms, all contribute to the greater likelihood that care will be sought for any particular symptom.

Although the Betas for several variables are statistically significant in the other three analyses, it should be observed that, with two exceptions, none of them amounts to 0.20. As a consequence, none of the variables that are statistically significant in these analyses could be described as contributing substantially to the prediction of the decision to seek medical care. The two exceptions are need variables.

These analyses were replicated using an index combining belief that the symptom was serious and belief that a physician would regard the symptom as serious, in place of the belief that a doctor could help relieve the symptom. Table 4 presents the percentage of individuals with symptoms who saw a doctor in each of the congruent and incongruent conditions. The overall pattern in Table 4 corresponds to that in Table 2. Aside from slight differences in percentages, the only real difference between Table 4 and Table 2 is that network advice seems to be somewhat more effective in precipitating a physician visit when the individual does not believe the condition is serious, than when the person does not believe a physician can help. On the other hand, about as high a percentage of people who feel the symptom is serious seek medical attention without network advice as do those who believe the doctor can do them some good.

TABLE 4. Per Cent of Respondents Initiating a First Doctor Visit for a Symptom by Network Advice and Perceived Seriousness of Symptom
(n = number of respondents in each category)

Proportion of Network Members Advising Respondent to See Doctor for Symptom	Symptom Serious	Symptom Not Serious
Some or all	85.6% (259)	49.3% (74)
None	65.2% (383)	13.4% (446)

Table 5 presents hierarchic multiple regression analyses for each of the congruent and incongruent conditions similar to those presented in Table 3. Again, the pattern and magnitudes of the total amounts of variance explained in these four analyses are quite similar to that found in Table 3. Thus, while less than 15 per cent of the variance is explained in the congruent conditions and in the condition when the symptom is believed to be serious but there is no network advice to go, 43 per cent of the variance is explained in the condition when the symptom is not believed to be serious but the network advises seeking medical attention. In the present analysis, substantial amounts of the surplus variance explained in this condition are accounted for with the sociodemographics and network characteristics, and the general health orientation items. Thus, in addition to income, marital

TABLE 5. Standardized Regression Coefficients for the Decision to Seek Medical Care for Symptoms by Network Advice and Perceived Seriousness of the Symptom

Step	Received Advice/ Serious (n = 259)	No Advice/ Not Serious (n = 446)	No Advice/ Serious (n = 383)	Received Advice/ Not Serious (n = 74)
I. Indicators of need				
Number of chronic conditions	0.06	0.11*	0.06	−0.07
Disability from symptom	0.13*	0.12†	0.09	−0.30‡
Subjective health status	0.17*	0.03	0.03	0.06
r^2 for Step I	0.03	0.03	0.02	0.09
II. Sociodemographics				
Total family income	−0.13	0.06	0.11	0.26†
Socioeconomic index	0.07	−0.07	−0.07	0.12
Race/ethnicity (white vs. other)	0.08	0.04	−0.02	0.00
Marital status (now married vs. other)	0.15*	−0.06	−0.08	−0.23‡
Sex (female = 1, male = 0)	−0.02	−0.09	−0.04	0.18
Age	0.09	0.08	0.18	0.37†
r^2 for Step II	0.07	0.05	0.05	0.28
III. Organization of care				
Continuity of care	0.17	0.07	0.10	0.09
r^2 for Step III	0.09	0.06	0.06	0.29
IV. Social network characteristics and health orientations				
Amount of network contact	0.02	−0.09	−0.14	−0.22
Amount of network consulting	−0.03	−0.06	−0.15*	−0.20
Network support	0.03	0.09	0.18*	0.36*
Network size—total number of members	0.06	0.06	−0.06	−0.23
Network advice to see doctor	−0.11	−0.10*	0.03	0.05
Acceptability	0.01	−0.03	0.01	−0.03
Accessibility	0.07	0.01	−0.01	0.03
Seriousness of illness	0.07	−0.11	−0.07	0.15
Perceived efficacy of care	−0.04	−0.03	−0.09	−0.01
Susceptibility to illness	−0.01	−0.01	0.05	0.02
Motivation—concern with health	0.05	0.05	0.03	0.02
Concern with medical costs	0.01	−0.06	−0.10	0.28‡
Tendency to delay seeking care	0.08	−0.09	−0.03	0.25‡
Total r^2 =	0.13	0.10	0.11	0.43
n	(259)	(446)	(383)	(74)

* $p < 0.05$.
† $p < 0.01$.
‡ $p < 0.10$.

status and age were also factors, the unmarried and older people being more likely to seek medical care on the advice of their networks but in opposition to their beliefs. Further, those who described their network as supportive, who were concerned about the costs of care, and who described themselves as delaying as long as possible before seeing a physician about an illness were more likely to seek medical attention in this condition. It also should be noted that, while disability was positively related to seeking medical attention when the individual did not believe a doctor could help, it is negatively related to seeking care when the person believes the condition is not serious. This suggests that restriction of activity may take place instead of a doctor visit when it is believed that the condition is not serious, but that activity restriction may precipitate a doctor visit when one's social network contradicts one's belief that a doctor can be of no help. Finally, it can be seen that although there are a number of statistically significant Betas in the regression analyses of the other three conditions, none of them achieves 0.20. Thus, none of the variables in these analyses are strongly related to the decision to seek medical care for symptoms in either of the congruent conditions nor are they strongly related to the decision to seek medical care when the individual believes the condition is serious but has not been advised to seek medical attention by his social network.

Conclusions

The first conclusion we drew from Table 1 is that the general characteristics of social networks and general health orientations have little effect on this decision. The present analyses suggest that these variables may influence the decision to seek medical care when they occur in interaction with the incongruent pattern of network advice to seek medical care and beliefs that are contradictory of that advice. Thus, the contribution of these variables to the amount of variance explained is 0.27 when the individual does not believe a doctor can help relieve the symptom and 0.14 when he believes the symptom is not serious. Because these variables make no appreciable contribution to the other analyses, these findings indicate that more general considerations are most likely to affect behavior when the opinions of members of a person's social network contradict what the person believes. Under these circumstances, behavior appears to be more amenable to the influences of general cognitive orientations. Indeed, this finding indicates that more enduring general factors may come into play in shaping behavior only when there is a conflict in the more specific and immediate factors relevant to the behavior. That similar influences did not appear in the other incongruent condition is explained by the fact that many members of the network may simply not advise the person on whether or not to seek care. Thus, for many of these people there may have been no conflict.

The extent to which retrospective reconstruction might be a problem also could not be determined from the data presented in Table 1. The analyses of congruent and incongruent patterns of network advice and personal beliefs, however, are at a minimum, supportive of a straightforward interpretation of Table 1. Thus, the two incongruent conditions, those in which retrospective reconstruction, to the extent that it occurred, did not lead to a consistent recall of the past, displayed patterns similar to what one might have expected were the reports of beliefs and network advice taken at face value. In the first incongruent condition, about two-thirds of those whose beliefs supported seeking medical care did so without network advice to seek attention. Nevertheless, when both beliefs and network advice support seeking care, a substantially larger percentage of the symptomatic go to the doctor. Thus, social network advice reinforces the probability that a behavior will occur as one might predict from knowledge about the beliefs related to that behavior.

In the second incongruent condition, when the network advises seeking care but beliefs do not support this behavior, fewer individuals actually seek care than in the other incongruent condition. Further, those individuals who do not believe the doctor can do much to relieve the symptom are less likely to respond to network advice to go to see the doctor than those individuals who do not believe the symptom is serious. This is precisely what one would expect, based on the social psychological tenet that a specific behavior is more readily predicted by beliefs that are directly related to that behavior than by beliefs that are tangential.[3] In this instance, the effects of network advice are greater when the belief with which the advice is incongruent is related less directly to the behavior being influenced. That is, individuals are more readily persuaded by networks to see a doctor when they believe the symptom is not serious than when they believe the doctor cannot help.

These findings appear to support the argument that self-reports regarding social network advice and personal beliefs about symptoms are valid. Indeed, if the self-reports were the result of retrospective reconstruction one would expect respondents to recall only congruent patterns of advice and beliefs. However, a large proportion of the respondents reported incongruent patterns. The conclusion that these self-reports of beliefs and advice regarding the symptoms are not the result of retrospective reconstruction, tentative though it is, has implications for research into the role of personal beliefs and network influences in the precipitation of decisions in a free-living population. As previously noted, it is extremely unlikely that prospective studies of such decision making will be possible in most circumstances. A prospective study of the decision to seek medical care for symptoms would require data collection from the moment the symptom is first experienced until the time when the decision regarding a visit to the doctor has been taken. Obviously, it simply is not feasible to collect data in this manner. The present evidence suggests, however, that data collected retrospectively may validly reflect the impact of both beliefs and network influences on the decision-making process.

References

1. Berkanovic E, Telesky C, Reeder S. Structural and social psychological factors in the decision to seek medical care for symptoms. Med Care, 1981, 19:693.

2. Sumner J. The 1976 LAMAS frame and master sample: technical description. Los Angeles: Institute for Social Science Research, 1976.

3. Fishbein M, Ajzen I. Belief, attitude, intention and behavior. Reading, Mass.: Addison Wesley, 1975.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Assessing the Effectiveness of Neonatal Intensive Care

PETER P. BUDETTI, M.D., J.D.,* AND PEGGY MCMANUS, M.H.S.†

Although neonatal intensive care has been praised widely for individual successes, its effectiveness has not been established systematically or conclusively. The literature consists principally of reports from individual intensive care units with sample sizes too small for statistical validation or generalization. This study analyzes the results obtained by recalculating and pooling the isolated reports. In addition, the findings of the solitary clinical trial, some scattered epidemiologic data, and several analyses of regional birthweight-specific time series data are reviewed. Taken together, this constitutes a body of evidence that supports the conclusion that intensive neonatal medical care has played a significant role in bringing about the impressive reduction in infant mortality that has taken place in this country since 1965.

NEONATAL INTENSIVE CARE is one of the factors often credited with the recent decline in infant mortality in the United States. The widespread diffusion and growing costs of newborn intensive care, together with concerns about the health of survivors and the potentially dehumanizing way such care may be delivered, however, have resulted in increased critical scrutiny of the benefits of neonatal intensive care. Policymakers, health professionals, and the public are raising serious questions about the costs of and outcome produced by highly technologic care for newborns. The specific issue to be resolved is whether or not increasing expenditures for newborn intensive care have resulted in measurable improvements in the health status of newborns.

This paper addresses what may be considered the principal indicator of the effectiveness of neonatal intensive care: whether premature and ill newborn infants do in fact have a better chance of survival with intensive care than without it. Randomized clinical trials would provide the most direct evidence of effectiveness, but few have been conducted. Instead the literature principally consists of epidemiologic surveys, reports from individual intensive care services, and birthweight-adjusted mortality rate analyses. Nearly all such reports make strong statements on behalf of medical care based on the remarkable declines in mortality during the years when neonatal intensive care proliferated.

Given the type of data available, however, some caution is necessary in inter-

---

* Institute for Health Policy Studies and Department of Pediatrics, School of Medicine, University of California, San Francisco, California.

† Institute for Health Policy Studies, School of Medicine, University of California, San Francisco, California.

Supported by Grant HS02975 from the National Center for Health Services Research. Based upon work done for the Congressional Office of Technology Assessment under Contract 933-22-60.0.

Reprint requests: Peter P. Budetti, Institute for Health Policy Studies, 1326 Third Avenue, San Francisco, CA 94143.

0025-7079/82/1000/1027/$01.45 © J. B. Lippincott Co.

pretation. Improvements in survival concurrent with advancements in intensive newborn care do not provide the information needed to determine the degree to which the gains in survival actually are due to intensive medical care of the newborn. Other factors such as lower prematurity rates or improved maternal health, nutrition, and intensive intrapartum care all undoubtedly have played a role.

The purpose of this article is to analyze the diverse and unsystematic literature as a step toward partitioning out the effect of neonatal care from that of other factors in determining newborn survival. This analysis addresses only survival and not the many questions about the functional status of survivors. Even this limited question is not resolved easily. On balance the available evidence does support the conclusion that neonatal intensive care has played a major role in improving the chances of survival of many newborns, particularly those of very low birthweight, but it remains impossible to quantify precisely how much of the improvement is due to medical care.

History of Neonatal Intensive Care and Infant Mortality in the United States

A historic perspective is needed to understand the debate over the effectiveness of modern methods of caring for newborn infants. In the early part of this century, most sick newborns died within the first few hours of life. Premature newborns were not expected to live more than a few days. In 1878, Dr. E. S. Tarnier instituted the use of a "warming chamber"—the first incubator—and was able to increase the chance of survival for a large number of premature newborns.[1] Similar machinery was widely used in international expositions, and later became common at amusement park sideshows featuring "incubator babies." Eventually incubators gained acceptance in the medical community and became standard equipment in most hospitals that had maternity services. As the survival period for premature infants lengthened, problems of nutrition and disease, to which such infants are prone, became evident and research on these problems began.

Facilities and medical techniques for care of the newborn have progressed rapidly since 1965 with the evolution of perinatal medicine and the development of associated medical technology. Care provided the newborn has passed through many phases and now involves the use of highly technologic, diagnostic, and therapeutic techniques and life-support systems. Along with changes in neonatal intensive care came changes in training physicians and nurses who deliver such care. Before 1960 responsibility for the care of the newborn was shared among general pediatricians, obstetricians, and general practitioners. In the early to mid-1960s, pediatricians with special interest in caring for sick newborns began to specialize in early infant care, leading to the development of neonatology as a pediatric subspecialty.[2] Now nurse clinicians and nurse specialists with advanced training and skills are responsible for delivering much of the intensive newborn care provided.

Neonatal intensive care typically is provided in organized hospital units classified into three groups or levels, depending on the sophistication and scope of the services they are equipped and staffed to provide. The distinctions among the three levels blur considerably in the field, however, and the resulting definitional confusions substantially complicate collecting and analyzing data.[3] Generally the results cited in this article are from nurseries that would qualify as Level III, the most sophisticated, although the level and criteria for classification often are not reported.

Remarkable improvements in infant mortality (deaths during the first year of life) have occurred in this country since 1915.[4–7] Most of this century's decline in infant mortality had taken place by 1950, however, and cannot be attributed to neonatal intensive care. The effects, if any, of newborn medical care only can be measured in terms of additional improvements in newborn survival subsequent to the introduction of modern methods. Between 1950 and 1965 the infant mortality rate continued to decline, but at a slower rate. During that period, nearly 60 per cent of the decline in infant mortality was experienced during the neonatal period (the first 28 days). This improvement in early survival is precisely the effect that would be expected if intensive newborn care was a major factor in improved survival. The end of this era coincides with the introduction of neonatal intensive care units (NICUs) and neonatology as a pediatric subspecialty.

Between 1965 and 1980 infant mortality fell by nearly 50 per cent, declining at 3.3 per cent annually, compared with the 1 per cent per year fall during the 1950–1965 period. Nearly all (88 per cent) of the decline since 1970 has occurred in the neonatal period, again consistent with a large medical care effect. Because the most recent improvements in the neonatal mortality rate were concurrent with the proliferation of intensive care newborn medical services, a number of studies have been undertaken to see whether the changes are, in fact, due to modern medical care.

The evidence collected falls into four categories: clinical trials, epidemiologic surveys, reports from individual intensive care units, and birthweight-adjusted mortality rate analyses. In reviewing this evidence, it is important to remember that each category addresses somewhat different aspects of intensive newborn care. Clinical trials of individual therapies are designed to test the efficacy of those measures in idealized situations. Clinical trials of a set of intensive measures compared with routine practice more closely would estimate the degree of effectiveness to be expected in general hospitals of the highest caliber. The reports from individual nurseries are a crude approximation of this same index of effectiveness. Epidemiologic surveys measure the net result on a community, including the practical issues related to access as well as the effectiveness of all available medical measures. By using an indicator that is related more directly to medical care than crude mortality rates, the birthweight-specific analysis provides a refinement on the epidemiologic surveys. Because the aspects of intensive newborn care assessed differ within each category they are presented separately.

Analysis of the Evidence on Effectiveness

Clinical Trials

There has been only one randomized controlled clinical trial of intensive care versus routine care of neonates.[8] That study, done in Australia in 1966–1970, showed improved survival for selected infants with very low birthweights, particularly those with respiratory distress syndrome (RDS), as experience in delivering intensive care developed. Kitchen and colleagues studied 238 of the 273 infants weighing between 1,000 and 1,500 grams who were born alive in the Royal Women's Hospital in Melbourne during that 4-year period. After excluding the 35 babies considered unsuitable for randomization, the authors assigned alternate patients to routine or intensive care. Intensive therapy consisted principally of umbilical artery cannulation for fluid and alkalai therapy and estimation of blood gases and acid-base status. Phototherapy, peripheral artery blood sampling, intermittent positive pressure ventilation (IPPV), and continuous positive airway pressure (CPAP) were not used for any infant.

Kitchen's results suggest that staff experience and newborn diagnosis are critical factors in determining outcomes. Improved survival was experienced by intensively treated patients throughout the study, but was statistically significant only during the latter 2 years. Even without the use of IPPV and CPAP, infants with RDS experienced significantly better survival with intensive care than with routine care in the latter period.

Had controlled trials been conducted in other nurseries during the late 1960s and early 1970s, there might be little question about the effect of intensive care methods on the survival of infants with very low birthweights. No other such studies have been identified, however. One leading neonatologist attributes this paucity of controlled studies to the notion that "controlled clinical trials of perinatal intensive care and regionalization have been neither feasible nor morally acceptable."[9] There have been clinical trials of some individual intensive care therapeutic techniques. One such study demonstrated improved mortality and shorter periods of hospitalization for infants who were transported under the care of a specialized neonatal team compared with those transported by the local ambulance authority.[10] The spectrum of specific intensive care measures studied in randomized clinical trials was reviewed recently by Sinclair and others.[11]

The debate over the efficacy of CPAP devices in the treatment of respiratory distress syndrome illustrates the status of most of the clinical evidence. This therapy has been hailed as a breakthrough, responsible for marked improvement in the survival of infants with RDS. However, the first controlled clinical trial was not conducted until several years after widespread adoption of CPAP and that study was not able to document statistically significant improvement in mortality.[12] Publication of that report in the pediatric literature was accompanied by a long editorial that reiterated the abundant evidence for the effectiveness of CPAP and detailed possible deficiencies of the controlled study, while expressing general support for clinical trials.[13] In short, many newborn therapies are unproven and even controversial. Thus no overall conclusions can be reached from the clinical literature alone.

Epidemiologic Surveys

Epidemiologic evidence suggests that medical care has helped lower mortality rates, but for the reasons discussed below the evidence is not conclusive. The Canadian experience showed a 50 per cent decline in neonatal mortality in Toronto in the decade following the organization of special neonatal care facilities.[14] An analysis of data for the United States between 1955 and 1973 demonstrated that, in general, mortality rates fell with increasing urbanization,[15] but urban areas were not stratified in terms of levels of intensive care available for newborns. Thus no conclusions could be drawn about the relationship between urban medical care and mortality.

In a major review article, Thompson and Reynolds summarized reports from various parts of the United States and Canada, all of which showed declines of 23 to 42 per cent after the development of regional intensive care facilities.[16] Although these data suggest that development of regional intensive care centers improved the outlook for infants in surrounding areas, the studies do not consider whether prematurity rates fell during the same period. Since declines in prematurity can markedly reduce neonatal mortality rates independent of changes in medical care, the omission greatly reduces the conclusions that can be drawn from these reports.

Sequential Reports From Individual Intensive Care Nurseries

Most of the literature on the effectiveness of neonatal intensive care consists of reports detailing the experience of individual nurseries. Twelve of these reports were reviewed by Thompson and Reynolds; all demonstrated reductions in inborn neonatal mortality, ranging from 17 to 76 per cent after the organization and designation of an intensive care newborn nursery.[16]

The impact that changes in prematurity rates can have on a nursery's mortality experience limits the value of these reports, however. If the distribution of births among birthweight groups changes from year to year, so that very small infants represent a lower proportion of total births, mortality experience will fall substantially. The proportion of infants in the very low birthweight category (1,500 grams or less) has the greatest impact on mortality rates.[17] Although those infants only represent about 1 per cent of all births, they account for approximately one half of all neonatal deaths.[3] Any individual nursery will have so few births in this category that even modest fluctuations in the proportion of infants with very low birthweights will have a marked impact on neonatal mortality.

The experience of the Bronx Municipal Hospital Center illustrates this phenomenon clearly.[18] In that facility, between 1966–1971 and 1972–1973, the in-hospital neonatal mortality rate declined from 16.9 per 1,000 live births to 11.2 per 1,000 live births. Over 75 per cent of that decline could be attributed to changes in the proportion of infants with very low birthweights (1,500 grams or less). That is, the neonatal mortality rate in the latter period would have been 12.6 per 1,000 live births just on the basis of a changed distribution of infants among birthweight categories.

The experience of the Bronx Municipal Hospital demonstrates that even in the absence of improvements due to medical care after birth, impressive declines in neonatal mortality can occur if a smaller proportion of high-risk infants is being born. Because the proportion of low birthweight infants (2,500 grams or less) declined by over 14 per cent between 1965 and 1978 in the United States,[19] reports of improvements in mortality must specify birthweight distribution if they are to help determine the effectiveness of newborn medical care. On the other hand, if mortality rates within specific birthweight groups do improve over time, this would be a strong indication that neonatal intensive care is having a beneficial impact. As Lee et al. have stated, if infants of 1,500 grams birthweight or less have a better chance of surviving after introduction of neonatal units, such a change must be due to "either improved neonatal medical care or to improved health status of the infants at birth, or both."[18]

A number of nurseries have reported mortality rates by birthweight before and after the introduction of an intensive care facility.[9, 20–26] The results for every nursery but one[26] show impressive declines in birthweight-specific mortality over time. Attributing most of these improvements to newborn medical care, Thompson and Reynolds felt that "NICUs have more than justified their reason for existence, reducing by approximately one-half the risk of neonatal death."[16]

Even birthweight-specific reports from individual nurseries are of limited usefulness, however, because of their small population samples. Often, only 20–30 infants in the very low birthweight groups are born during the study period, making it difficult to determine whether improvements over time are statistically significant. For this reason we grouped all available reports of mortality rates for inborn infants into 5-year periods beginning in 1961, recalculating published figures into comparable categories as necessary. These pooled data demonstrate significant declines over time for infants with very low birthweights (Table 1 & 2, Figure 1).Mortality has fallen from approximately 50 per cent to less than 20 per cent for infants with birthweights between 1,001 and 1,500 grams, while the extremely small newborns (1,000 grams or less) have improved from over 93 per cent mortality to just over 50 per cent.

These impressive trends over time demonstrated by the pooled birthweight-specific mortality data strongly suggest that

TABLE 1. Pooled Neonatal Mortality Data for Inborn Infants Weighing 1,000 Grams or Less

Hospital	1961–1965	1966–1970	1971–1975	1976–1980
University of California, San Francisco[22]	28/29	26/34	22/33	15/30
Children's, San Francisco[21]			30/35	8/24
University of Washington[29,30]		27/29		25/41
Medical Center, Columbus, Ga.[25]	157/168	27/34	26/35	
Bronx Municipal[18]		119/138	18/25	
University of Illinois[40]			29/38	
Harvard[41]				82/146
Cambridge Maternity[23]				9/25
Royal Victoria[9]		65/73	58/75	
Simpson Memorial, Edinburgh[42]		84/92		
University College Hospital, London[43]		29/36	26/33	
East Hospital, Göteborg, Sweden[24]		4/7		
Hammersmith Hospital, London[26]	37/41	40/46	20/23	
Total*	222/238 (933)	421/489 (861) †	229/297 (771) ‡	139/266 (523) §

Rates for each 5-year period differ significantly from those of each previous 5-year period.

* Numbers in parentheses are neonatal mortality rates (deaths/1,000 live births).

† $X^2c = 7.40$ ($p < 0.01$).

‡ $X^2c = 9.82$ ($p < 0.005$).

§ $X^2c = 37.18$ ($p < 0.001$).

intensive newborn medical care has helped to improve survival for infants with very low birthweights. All available series were included and it seems reasonable to assume that the range of evolving intensive therapies applied over each 5-year period was roughly comparable in the reporting nurseries. Thus, these pooled data reflect average results over time in intensive care nurseries. Nevertheless, these trends are based on nonexperimental, historic data and as such cannot be regarded as conclusive.

Comparison of the pooled data with the results reported by Kitchen et al. from Australia emphasizes the importance of clinical trials with contemporaneous controls.[8] Including the cases excluded from their study, Kitchen's group experienced a rate of 399 deaths per 1,000 live births for the entire population of infants in the 1,000–1,500 gram category born in their hospital. That result is comparable to the average rate of 372 deaths per 1,000 live births reported for intensive care by other nurseries during the same period, 1966–1970 (Table 2). Thus, during the years of a clinical trial, the experience of the Australian nursery as a whole was as good as that of other centers, even though it included infants who were

TABLE 2. Pooled Neonatal Mortality Data for Inborn Infants Weighing 1,001–1,500 Grams

Hospital	1961–1965 Deaths/Births	1966–1970 Deaths/Births	1971–1975 Deaths/Births	1976–1980 Deaths/Births
University of California, San Francisco[22]	26/55	23/63	15/80	9/68
Cambridge Maternity[23]				23/96
Royal Victoria[9]		25/89	14/90	
Harvard[41]				37/213
University of Washington[39]				14/75
Medical Center, Columbus, Ga.[25]	116/219	14/44	14/41	
Bronx Municipal[18]		69/193	11/42	
Simpson Memorial, Edinburgh[42]		71/154		
East Hospital, Göteborg, Sweden[24]		10/24		
Hammersmith Hospital, London[26]	38/85	40/81	34/81	
Royal Women's Hospital, Melbourne[44]	135/277	82/250		
Total*	315/636 (495)	334/898 (372)†	88/334 (263)‡	83/452 (184)§

Rates for each 5-year period differ significantly from those for each previous 5-year period.
* Numbers in parentheses are neonatal mortality rates (deaths/1,000 live births).
† $X^2c = 22.7$ ($p < 0.001$).
‡ $X^2c = 12.24$ ($p < 0.001$).
§ $X^2c = 6.73$ ($p < 0.01$).

allocated to routine care by the randomization process.

The controlled study design allowed Kitchen et al. to demonstrate the marginal value of applying specific additional intensive resources to infants with very low birthweights. During the last 2 years of that study, more than one third of the routinely treated group did not survive, compared with less than one fifth of the study infants. Thus, these figures allow a reasonably precise estimation of the value of the additional care, in contrast to the pooled nursery data. While the pooled data strongly suggest that the general improvements in survival over time represent the cumulative effect of new therapeutic modalities, they do not provide the direct evidence necessary to form such a conclusion with confidence.

Birthweight-Adjusted Mortality Rate Analyses

It is not possible to compare the experience in individual nurseries with the survival rates of infants of similar birthweight who were not cared for in intensive newborn units. As pointed out above, it is difficult to identify the levels of care available at different hospitals confidently. Moreover, not all states link birth certificates with death certificates, so it is impossible to know on a routine basis the birth-

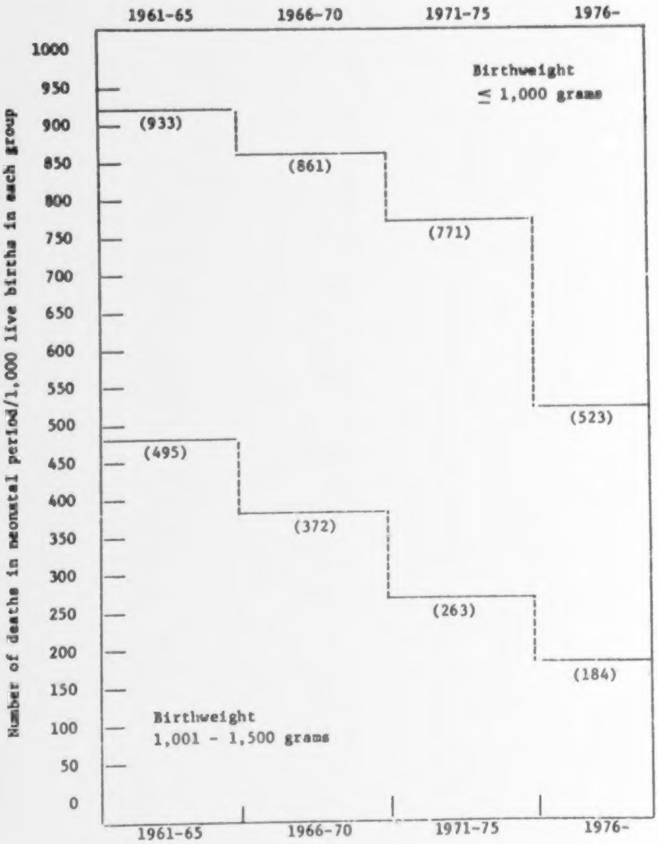


FIG. 1. Pooled neonatal mortality data.

weight of individual infants who die. Linked data for several areas have now been reported, however, and show major improvements in neonatal mortality between 1960 and the present.

Ronald L. Williams of the University of California, Santa Barbara, has analyzed changes in mortality rates among newborns in California over the past two decades.[27-30] Between 1960 and 1977 the neonatal mortality rate in California fell by 57 per cent. During the same period, low birthweight rates also declined by some 15 per cent. It was necessary, therefore, to partition the improvement in neonatal mortality into the part that was due to changes in birthweight distribution and the part due to other factors, including medical care. In contrast with the result noted above for Bronx Municipal Hospital Center, Williams showed that nearly 85 per cent of the remarkable decline in neonatal mortality in California resulted from birthweight-specific declines.[30] That is, the reduction in prematurity only accounted for about 15 per cent of the decline in neonatal mortality, while medical care and/or better health status of individual infants accounted for the remainder. More than half of the overall decline could be attributed to improvements in the mortality rates of infants with low birthweights. Williams also found that blacks experienced lower mortality rates than whites within low birthweight groups. This partially offset the fact that the birthweight distribution among blacks actually became less favorable over the 1960–1977 period.

Kleinman and others at the National Center for Health Statistics analyzed data from six states and reached similar conclusions.[31] They found that weight-specific changes accounted for nearly 80 per cent of the decline in early neonatal mortality rates. More than half of the overall decline resulted from improvements in mortality among the low birthweight groups. They estimated that, for the nation as a whole, about 83 per cent of the decline in early neonatal mortality for whites and 97 per cent of the decline for nonwhites could be attributed to factors other than changes in birthweight.

The same investigators discussed factors that could be causing the decline in birthweight-specific mortality rates. They rejected a suggestion that lower death rates were merely an artifact resulting from improved registration. They also concluded it was unlikely that any major shift in distinctions between live births and fetal deaths (stillbirths) had taken place, since fetal deaths were also declining. Finally, they attempted to deal with the question of whether infants within given birthweight groups might have different risk characteristics over time. They found that comparing gestational age distributions made no difference for the calculation and they concluded:

> . . . The most likely explanation for at least part of the recent decline in mortality levels among low birthweight infants is the development of the medical technology for the successful management of premature infants and the consequent proliferation of this technology. While there are many components of the technology which have not been critically evaluated, there are indications of effectiveness.[31]

Lee et al. examined the effect of birthweight changes on the improvements in neonatal mortality in the United States between 1950 and 1975. They determined that no improvement had taken place in the weight distribution or, with the exception of perinatal medical care, other factors known to affect survival.[32] Their conclusion, similar to those of Williams and Kleinman was that:

> [t]he most plausible explanation . . . is the steady improvement in perinatal medical care which has made for greater infant survival at a given birthweight.

Most recently, these same investigators used birthweight-adjusted analysis to examine variations in neonatal mortality among the 50 states and the District of Columbia, and among 13 industrialized nations.[17] They confirmed the role of the very low birthweight rate in predicting neonatal mortality; that rate accounted for about three quarters of the variance among the populations they studied. They suggested that variance of neonatal mortality after holding constant the very low birthweight rate might be a useful preliminary indicator of the quality of newborn care.

Discussion

Existing studies are helpful in answering part of the question regarding the role of medical care in bringing about improvements in infant mortality, but they do not settle all the issues. Of particular value are the birthweight-adjusted mortality rate analyses. By estimating that only 15 per cent of the decline in overall mortality rates is simply the result of changes in birthweight distribution, these analyses define the maximum impact on survival that could be attributed to medical care of the newborn. However, this is not to say that all of the remaining 85 per cent is necessarily a result of neonatal medical care.

One confounding problem would be any underlying changes in the health status of infants within each weight group that improved neonatal mortality independent of intensive newborn care. Such changes could be the result of medical or nonmedical factors. Medical factors, such as intensive intrapartum care or steroid therapy during premature labor to prevent hyaline membrane disease (HMD), could produce infants that are more likely to survive in the extremely fragile low birthweight groups. Obstetric care during pregnancy prior to labor would not be expected to affect birthweight-specific mortality rates but to reduce the chances of low birthweight.

Nonmedical factors such as improved maternal nutrition and socioeconomic status also have their primary beneficial effect through higher birthweights.[33,34] It is possible to speculate that better maternal nutrition and general health also result in subtle biochemical or other changes that further increase the chance of survival for infants in each weight group. Unfortunately, none of these hypotheses can be established or rejected confidently from the available data.

Another factor that complicates interpretation of the data is the learning curve associated with new technologies. The maximal impact of innovations may not appear for several years until staff expertise has developed. Similarly, favorable results in applying a new therapy to infants of a certain size may encourage its use in ever smaller infants, thus requiring additional time to see the maximal effect. Kitchen et al. noted that survival for routinely treated patients had improved over time even though all such infants were "managed in apparently the same fashion . . . "[8] They cautioned that this finding "illustrates the dangers inherent in drawing conclusions from retrospective or sequential data."

Similarly, one sketchy report describes a British hospital's experience in managing ill and premature babies with little or no intensive care.[35] No respirators, phototherapy, continuous positive airway pressure (CPAP) device, nor extensive mask and glove isolation were used in the nursery. Some modern techniques, including initial resuscitation, catheterization of umbilical vessels, exchange transfusions, and incubators were applied. The absolute levels of mortality achieved in each weight category would have been unacceptable at a major nursery in the United States during the same time period; however, there was significant annual improvement, even within birthweight groups. On the basis of this limited report, one is not able to conclude whether this nursery is, in fact, able to improve mortality without changing neonatal care and without delivering the most advanced methods. Such reports do serve to illustrate the continuing need for new evidence to determine the precise contribution of neonatal care in improving birthweight-specific mortality rates.

The work by Williams provides some further indirect evidence that supports the hypothesis that medical care was a major factor in lowering mortality rates. For infants in all hospitals in California, he found that once the risks faced at each hospital are taken into account, hospitals with neonatal intensive care nurseries have better survival rates than those without intensive care units, and hospitals with large delivery services and high specialist–generalist ratios have the best performances.[28,36]

In summary, while it may not be possible to determine the precise role that neonatal intensive care has played in lowering mortality rates among low birthweight infants, the evidence supports the conclusion that neonatal care has been a significant factor. The degree, if any, to which infants are somehow improving their inherent ability to survive from year to year and the role of intensive intrapartum obstetrical care could be tested only through additional controlled clinical trials.

Conclusion

Present knowledge provides only tentative conclusions on the effectiveness of neonatal intensive care. Nationally, infant mortality has fallen by an impressive 49 per cent (from 24.7 per 1,000 live births in 1965 to 12.5 per 1,000 live births in 1980) during the recent period when access to increasingly sophisticated care grew substantially. Among the factors that may have led to this dramatic decline are improved prenatal and maternity care, lower fertility rates and improved maternal nutrition, as well as improvements in the medical care provided to ill or extremely small infants. It has been difficult to assess the impact of medical care because of inadequate information on the actual types of neonatal intensive care provided and their relationship to medical outcome.

Within these limitations, partial answers to the questions of the value of medical care for newborns are available at present. These answers suggest that medical care of the newborn has been effective in many ways, but present information is inadequate to determine the degree to which the amount or type of medical care is associated with improved outcome. Much of the improvement in mortality rates for newborns in the United States has been in the areas where medical care would be expected to have its greatest impact: survival of infants with very low birthweights (less than or equal to 1,500 grams, or 3.3 pounds) has risen from about one fourth to over two thirds of infants of that size, and virtually all of the improvement in survival occurred during the first 28 days of life. The solitary clinical trial supports the crude epidemiologic data, the individual nursery success stories, and the encouraging birthweight-adjusted time series data. Together, this constitutes a body of highly suggestive evidence that intensive care of the newborn has had beneficial effects on survival.

The results of previous studies, then, tell us that more intensive care generally is associated with better survival, but they do not help us determine the critical elements of such care that make a difference. Existing studies treat intensity of medical care (defined as type and amount of services provided) as a "black box," failing to distinguish the roles of case-mix, staffing, treatment modes, size, utilization, and the other variables that comprise intensity. The present system of categorizing nurseries into different levels of intensity is not precise enough nor sufficiently standardized, so that great variation exists among nurseries. Consequently, the levels of intensity and the possible relationship between the intensity of care and outcomes cannot be specified in ways that would be useful for planning or making decisions about regionalization or about relating reimbursement to levels of care or other indicators of cost.

Resolution of the remaining questions will require a multi-tiered strategy. Studies such as the Australian controlled trial would violate current standards of care and probably can no longer be attempted for ethical and legal reasons. Individual unproven technologies, as well as future innovations, could and should be evaluated through prospective controlled clinical trials. Retrospective and current analyses testing the relationship between survival rates, and the diffusion of advanced personnel and equipment would be a valuable contribution. In particular, studies that analyze the relationship between specific newborn intensive care resources and health outcomes for selected conditions would provide new information on the impacts of specific organizational and technologic components of medical care. The high personal and financial costs of providing intensive newborn care require more definitive evidence concerning the effects of such care.

Finally, an equally important question left unanswered by survival figures is whether the increased number of infants being saved will be able to lead normal lives, or whether they will suffer from serious handicaps such as cerebral palsy and mental retardation. Other serious health problems of survivors may include high rates of child abuse; learning, language and behavior disorders; and complications of therapy such as blindness düe to retrolental fibroplasia and impaired hearing secondary to the use of certain antibiotics. These questions are perhaps even more perplexing than the issue of survival because the absolute number of afflicted individuals can increase at the same time that the incidence of individual problems is decreasing. Such a paradoxical result will occur if mortality rates fall faster than the rates of decline in the incidence of residual disorders. This aspect of intensive care is being addressed,[37,38] but is still in need of additional long range, intensive analysis.[3]

Acknowledgments

Special thanks are due to Nancy Barrand, LuAnn Heinen, and Suzanne Stenmark for their contributions, to Carol Bok, Barbara Johnson, and Les Gates for their preparation of the manuscript, and to our colleagues at the Institute for Health Policy Studies, particularly Steven A. Schroeder, for reviewing the earlier drafts.

References

1. Silverman WA. Incubator-baby sideshows. Pediatrics 1979;64:127.

2. Schlesinger ER. Neonatal intensive care: planning for services and outcomes following care. J Pediatr 1973;82:916.

3. Budetti P, McManus P, Barrand N, Heinen L. The costs and effectiveness of neonatal intensive care. Washington, D.C.: Congressional Office of Technology Assessment. U.S. Government Printing Office, 1981.

4. National Center for Health Statistics. Vital statistics of the United States 1975, vol 2 Mortality, part A. Hyattsville, Md.: National Center for Health Statistics, 1979. (DHEW Publication No. (PHS) 79-1114).

5. National Center for Health Statistics. Final mortality statistics 1977. Hyattsville, Md.: National Center for Health Statistics, 1979. (Monthly Vital Statistics Report, Advance Report. DHEW Publication No. (PHS) 79-1120, vol 28, No 1 Suppl.).

6. National Center for Health Statistics. Final mortality statistics 1978. Hyattsville, MD: National Center for Health Statistics. 1980. (Monthly Vital Statistics Report, Advance Report. DHHS Publication No. (PHS) 80-1120, vol 29 no 6 Suppl. 2).

7. National Center for Health Statistics. Births, marriages, divorces and deaths for 1980. Hyattsville, Md.: National Center for Health Statistics, 1980. (Monthly Vital Statistics Reports, Provisional Data. DHHS Publication No. (PHS) 81-1120, vol 29, no 12).

8. Kitchen WH, Ryan MM, Rickards A, et al. A longitudinal study of very low-birthweight infants. I. Study design and mortality rates. Develop Med Child Neurol 1978;20:065.

9. Usher R. Changing mortality rates with perinatal intensive care and regionalization. Seminars Perinatol 1977;1:309.

10. Chance GW, Matthew JD, Gash J, Williams G, Cunningham K. Neonatal transport: a controlled study of skilled assistance. J Pediatr 1978;93:662.

11. Sinclair JC, Torrance GW, Boyle MH, Horwood SP, Saigal S, Sackett DL. Evaluation of neonatal-intensive-care programs. N Engl J Med 1981;305:489.

12. Belenky DA, Orr RJ, Woodrum DE, Hodson A. Is continuous transpulmonary pressure better than conventional respiratory management of hyaline membrane disease? A controlled study. Pediatrics 1976;58:800.

13. Wung JT, Stark RI, Hegyi T, Driscoll J, James LS. CDP: a major breakthrough! Pediatrics 1976;58:783.

14. Swyer PR, Goodwin JW, eds. Regional services in reproductive medicine. Toronto: The Joint Committee of the Society of Obstetricians and Gynecologists of Canada and the Canadian Pediatric Society, 1973.

15. Eisner V, Pratt NW, Hexter A, Chabot MJ, Sayal N. Improvement in infant and perinatal mortality in the United States, 1965–1973: I. Priorities for intervention. Am J Public Health 1978;68:359.

16. Thompson T, Reynolds J. The results of intensive care therapy for neonates. Perinatal Med 1977;5:59.

17. Lee KS, Paneth N, Gartner LM, Pearlman M. The very low birthweight rate: principal predictor of neonatal mortality in industrialized populations. J Pediatr 1980;97:759.

18. Lee KS, Tseng PI, Eidelman AI, Kandall SR, Gartner LM. Determinants of the neonatal mortality. Am J Dis Child 1976;130:842.

19. Bureau of the Census. Statistical abstract of the United States 1980. Washington, D.C.: U.S. Government Printing Office, 1980. (Table 91.)

20. Alden ER, Mandelkorn T, Woodrum DE, Wennberg R, Parks C, Hodson WA. Morbidity and mortality of infants weighing less than 1,000 grams in an intensive care nursery. Pediatrics 1972;50:40.

21. Hirata T, Halliday HL. Survival of infants of 1,000 g birth weight or less. (Letter), Br Med J 1977;2:1670.

22. Phibbs RH. Survival rates by birthweight. San Francisco: University of California Nurseries, 1979. (Unpublished).

23. Robertson NRC. Letter. Lancet 1979;ii:362.

24. Sabel KG, Olegard R, Victorin L. Remaining sequelae with modern perinatal care. Pediatrics 1976;57:652.

25. Thompson RE, Souma ML, Cassady G, Sumners J. Impact of intensive newborn care on neonatal mortality in a community hospital. South Med J 1976;69:688.

26. Jones RA, Cummins M, Davies PA. Infants of very low birthweight: a 15-year analysis. Lancet 1979;i:1332.

27. Williams RL. Intrauterine growth and neonatal risk in California. (Appendix) Cunningham GC, Hawes W, Madore C, eds. Santa Barbara: University of California Community and Organization Research Institute, 1976.

28. Williams RL. Measuring the effectiveness of perinatal medical care. Med Care 1979;17:95.

29. Williams RL, Cunningham GC, Norris FD, Tashiro M. Monitoring perinatal morality rates: California 1970 to 1976. Am J Obstet Gynecol 1980;136:559.

30. Williams RL, Chen P. Identifying the sources of the recent decline in perinatal mortality rates in California. N Engl J Med 1982;306:207.

31. Kleinman JC, Kovar MG, Feldman JJ, Young CA. A comparison of 1960 and 1973–1974 early neonatal mortality in selected states. Am J Epidemiol 1978;108:454.

32. Lee KS, Paneth N, Gartner LM, Pearlman M. Neonatal mortality: an analysis of the recent improvement in the United States. Am J Public Health 1980;70:15.

33. National Center for Health Statistics. Teenage childbearing: United States, 1966–75. Hyattsville, Md.: National Center for Health Statistics, 1977. (Monthly Vital Statistics Report, Natality Statistics. DHEW Publication No. (PHS) 77-1120, vol 26, no 5 Suppl.).

34. Taffel S. Factors associated with low birthweight, United States, 1976. Hyattsville, MD: National Center for Health Statistics, 1980. (Series 21, No. 37).

35. Hughes-Davies TH. Conservative care of the newborn baby. Arch Dis Child 1979;54:59.

36. Williams RL. Measuring the effectiveness of perinatal medical care. Santa Barbara: University of California Community and Organization Research Unit, 1980. (U.S. Bureau of Community Health Services Grant Report MC-R-060390-03).

37. Shapiro S, McCormick M, Starfield B, Krischer J, Bross D. Relevance of correlates of infant deaths for significant morbidity at 1 year of age. Am J Obstetr Gynecol 1980;136:363.

38. McCormick M, Shapiro S, Starfield BH. Rehospitaliziation in the first year of life for high-risk survivors. Pediatrics 1980;66:991.

39. Jackson NS. Unpublished data. Division of Neonatal Biology. Department of Pediatrics, University of Washington, 1979.

40. Bhat R, Raju TNK, Vidyasagar D. immediate and long-term outcome of infants less than 1,000 grams. Crit Care Med 1978;6:147.

41. Brown ER, Taeusch HW. Letter. Lancet 1979;ii:362.

42. Drillien, CM. Aetiology and outcome in low-birthweight infants. Dev Med Child Neurol 1972;14:563.

43. Stewart AL, Turcan D, Rawlings G, Reynolds EOR. Prognosis for infants weighing 1,000 g or less at birth. Arch Dis Child 1977;52:97.

44. Kitchen W, Campbell D. Controlled trial of intensive care for very low birthweight infants. Pediatrics 1971;48:711.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Improving Physician Compliance With Preventive Medicine Guidelines

DAVID I. COHEN, M.D., M.SC.,* BENJAMIN LITTENBERG, B.A.,†
CHERYL WETZEL, B.A.,‡ AND DUNCAN vB. NEUHAUSER, PH.D.§

Similar general medical outpatient clinics with randomly assigned patients were used to evaluate the effectiveness of a program that was to increase house staff compliance with preventive medicine guidelines. Two clinics were designated experimental and two served as controls. In the experimental clinics, age-specific checklists of all recommended preventive procedures (drawn from the Canadian Task Force report on The Periodic Health Examination and American Cancer Society guidelines) were appended to each patient's chart. In addition, house officers were presented with a series of weekly seminars dealing with issues in screening, as well as the specific recommendations included in the checklist. House officers in all four clinics were tested for their knowledge and attitudes toward the preventive program before and after the intervention. Counts of immunizations and mammograms performed and the total populations eligible for these procedures were determined for all four clinics. As predicted, test scores as well as mammography and immunization rates increased significantly (from 2–10 per cent) in the intervention clinics as compared with controls. We conclude that this intervention was clearly effective in the short run. However, follow-up studies will be necessary to determine whether the desired long-term effect has been achieved.

TOO LITTLE PREVENTIVE care is provided. What can be done to change the behavior of physicians responsible for providing such care? The academic medical center has been recognized as a hub for diffusing information and technology. It is an environment that fosters the acquisition of knowledge and the shaping of attitudes. A number of studies have examined the impact of various maneuvers upon house officers' use of laboratory testing. Several of these studies are summarized in the first of our "Metro Firm Trials" published in this journal.[1] Most studies of physician behavior change have explored the apparent overutilization of diagnostic testing in the teaching hospital setting. We, however, focused our attention upon the use of screening and preventive services, which have been underutilized by all indications.[2] A recent review by Carter, Belcher, and Inui[3] has chronicled some of the explanations for this underutilization, which include physician, patient, societal, and economic factors.

In this study, we test under controlled conditions and show the effectiveness of a simple intervention to improve house of-

---

* Assistant Professor, Department of Medicine, Case Western Reserve University, Chief, Division of General Medicine, Cleveland Metropolitan General Hospital, Cleveland, Ohio.

† Medical student, Case Western Reserve University, Cleveland, Ohio.

‡ Research assistant, Cleveland Metropolitan General Hospital, Cleveland, Ohio.

§ Professor, Department of Community Medicine and Epidemiology, Case Western Reserve University, Cleveland, Ohio.

Supported by Cleveland Foundation and Primary Care Funds, State of Ohio Board of Regents.

Portions of this article were presented at the Midwestern Regional Meeting of the American Federation of Clinical Research and at the National Meeting of The Robert Wood Johnson Foundation Clinical Scholars Program.

Reprint requests: David I. Cohen, M.D., M.Sc., Cleveland Metropolitan General Hospital, 3395 Scranton Road, Room H-331, Cleveland, OH 44109.

0025-7079/82/1000/1040/$01.55 © J. B. Lippincott Co.

ficer compliance with preventive medicine guidelines. We observe the relationship between physicians' knowledge and attitudes, and their use of preventive testing.

The Setting

The study was conducted in the Medical Outpatient Department at Cleveland Metropolitan General Hospital in the fall, 1980.

The Department of Medicine was reorganized during 1975 into a system composed of four general medical firms,[4] each consisting of a 28-bed inpatient unit as well as its own outpatient clinic. House officers are randomly assigned to a firm during internship and remain with that firm for the duration of their training. Patients similarly are randomized into one of the four firms, either on inpatient admission or on first contact with the outpatient area. They are assigned to a specific house officer with whom they remain associated until the house officer leaves the program. Then they are reassigned to another house officer in the same firm. The system thus has provided us with a measure of continuity not generally available in most academic medical centers.[5] Continuity is seen in the inpatient and outpatient services and has enhanced physician–patient relationships, clinic and ward team associations, and faculty–house staff teaching relationships. Moreover, the firm system has provided us with a unique laboratory for clinical and health care research. In essence what we have are four similar groups of patients and physicians, enabling us to study the effects of a maneuver in one or more films, while using the others as controls.

Methods

For this study two firms were randomly designated experimental and one served as a control. Three age-specific checklists, each delineating recommended preventive screening tests and procedures for all patients falling within a given age group, were designed. Items included in the lists were limited to those given Class A recommendations by the Canadian Task Force on the Periodic Health Examination[6] and those recommended by the American Cancer Society (Table 1).[7]

On the experimental firms the appropriate checklist (by age) was affixed to the chart cover of each patient who came to the medical clinic to serve as a reminder of the appropriate preventive measures for that patient. A trained research assistant quickly scanned each medical record to determine whether any of the recommendations had been followed in the recent past. Her findings were recorded on the checklists and appropriate radiology requests were completed as indicated. The house officer who was responsible for the

TABLE 1. Items Included on Preventive Medicine Checklists

	Age 16–44	Age 45–64	Age 65+
Required	Check blood pressure Pap test Breast exam Tetanus immunization	Check blood pressure Pap test Breast exam Mammogram (50–60) Stool hematest Tetanus immunization	Check blood pressure Stool hematest Pneumovax Influenza immunization Tetanus immunization

The following are optional for high risk groups at physicians' discretion: PPD, VDRL, GC culture.

primary care of the patient then was left to determine which procedures should be followed. Final orders for all screening tests and immunizations thus were made by the primary care physician.

The checklist intervention proceeded for a period of 4 months. Specific counts of the vaccinations and mammograms performed were kept by the clinic nursing staff and the radiology department for the two experimental firms and one control firm. The numbers of patients seen in the clinics who were eligible for these procedures were compiled by a research assistant. The percentages of eligible patients who actually received mammograms, influenza immunizations, and pneumovax during the period of the study were tabulated with these data. Furthermore, these data enabled us to identify previously screened patients to prevent duplications.

At the time the checklists were introduced into the medical clinics, a series of five seminars dealing with issues of screening and preventive medicine was offered to the medical house staff. These seminars were given weekly during the first 2 months of the clinic intervention. The first session was devoted to a general discussion of screening and of the methods employed to determine the efficacy of various screening maneuvers and preventive interventions. Subsequent sessions dealt with the specific recommendations being encouraged in the medical clinics. While house officers from all firms were invited to these sessions, those from the experimental firms were particularly encouraged to attend. In fact, with only one exception, no house officers from the control firms attended, and even attendance by members of the experimental firms was poor, averaging 1.2 seminars per physician for the 22 physicians in these firms.

A test instrument was developed to assess physicians' knowledge of and attitudes toward the use of preventive procedures and screening maneuvers. This test was administered to all medical house officers and faculty prior to the experimental intervention and seminar series, and again to all house officers at the conclusion of the study. The test consisted of two parts, one dealing with factual information drawn directly from the Report of the Canadian Task Force on the Periodic Health Examination and from the American Cancer Society guidelines, and the other dealing with attitudinal issues concerning physicians' choices of screening tests in varied clinical situations. Positive scores for the attitudinal section were given for those screening tests judged appropriate for a given situation by a panel of experts; negative scores were given for those judged inappropriate.

Results

The results of the intervention are summarized in Table 2. Over the 4-month period of the study only 5 per cent of the eligible controls received pneumovax, while 42 per cent of patients on the experimental firms were immunized. Influenza immunizations were delivered to 4 per cent of the eligible population on the control firm and 36 per cent of the experimental group. Similarly, mammography was performed in only 4 per cent of the control group and 32 per cent of eligible patients in the experimental firms. All differences are significant at the $p < .001$ level.

The data are graphically presented in terms of monthly rates in Figure 1. Virtually none of our patients routinely received screening mammograms or immunizations at the start of our intervention. This occurred partly because the influenza vaccine for that year was released just at the beginning of the study. The effects of the intervention, however, were seen immediately. Pneumovax and influenza immunization rates rose 30 per cent on the intervention firms during the first month of the study and continued to rise to almost 60 per cent during the subsequent 2 months,

TABLE 2. Immunization and Mammography Rates Among Eligible Patients During 4-Month Intervention

	Experimental Firms (8B & 11C)		Control Firms (6C)	
	Number Eligible	% Completed	Number Eligible	% Completed
Pneumovax	547	42%	291	5%
Influenza	581	36%	291	4%
Mammogram	290	32%	138	4%

$p < 0.001$ in all cases.

finally dropping during the fourth month. Rates on the control firm remained less than 10 per cent for the duration of the study. Similar results were seen in the case of mammography screening. It should be noted that it was impossible to eliminate previously screened patients from our tabulation of denominators, that is, the populations eligible for immunizations and mammograms. Therefore rates reflect the most conservative possible estimates of completed procedures. Furthermore, because patients tend to return to our medical clinics at about 3-month intervals, it is conceivable that the drop noted toward the end of the study represents the return of patients who had already undergone appropriate screening and vaccination for that year.

The data presented up to this point have been analyzed in terms of the patient population on the firms. Because our intervention was directed toward improving the delivery of preventive measures to our patients, we felt that the patient population was an appropriate unit of analysis to evaluate its effectiveness. However, our interest in the effect of our intervention upon physicians' attitudes and behaviors necessitated that the data also be examined using the physician as unit of analysis.

Differences between mean scores for delivery of mammography, influenza immunization, and pneumovax by physicians on the experimental and control firms all were significant at the $p < .001$ level. Similarly, point-to-serial correlations using Pearson's R showed highly significant differences between experimental and control groups in support of the experimental hypotheses. Membership on the experimental firms was correlated strongly with the delivery of mammography ($r = 0.53$, p

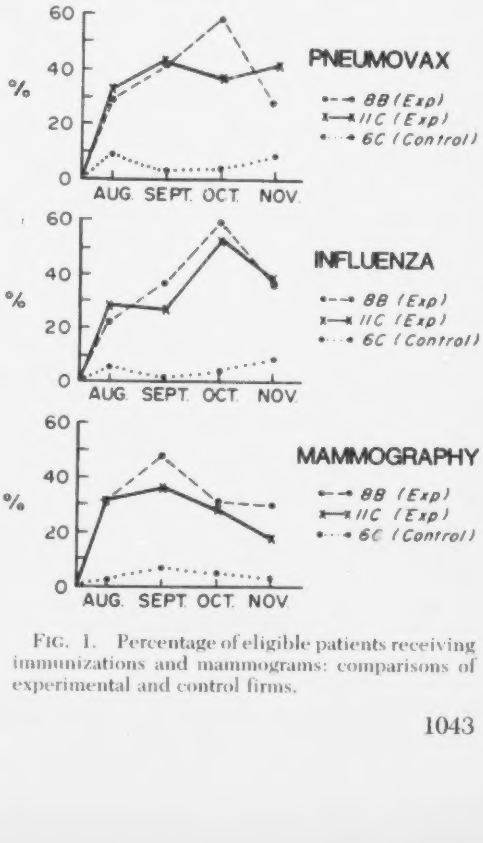


FIG. 1. Percentage of eligible patients receiving immunizations and mammograms: comparisons of experimental and control firms.

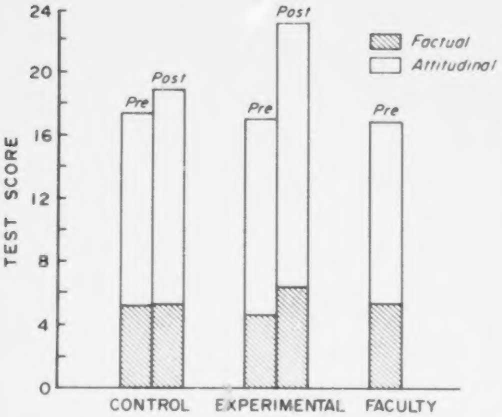


FIG. 2. Comparison of pretest and posttest scores between experimental and control firms.

< 0.001), influenza immunizations ($r = 0.65$, $p < 0.0001$), and pneumovax ($r = 0.70$, $p < 0.0001$).

Scores of the pretest and posttest were tabulated following the clinic intervention to assure that there was no bias in terms of faculty–house staff interactions. Three separate scores were calculated, one for the total test, a perfect score being 36; one for the attitudinal portion alone, a perfect score being 27; and one for the factual portion alone, a perfect score being 9. The results are summarized in Figure 2. No statistically significant differences were observed between the mean factual, attitudinal, or total scores of control and experimental groups on pretesting. Similarly, no differences were noted between house staff and faculty mean scores on pretesting. However, a significant difference was observed in the mean attitudinal and total test scores ($p < 0.05$ in both cases) on posttesting between the experimental and control groups. Similarly, the differences between experimental and control groups in the improvement in total scores from pretest to posttest was significant ($p < 0.05$). Unexpectedly, no difference was observed in the mean posttest factual scores between experimental and control groups.

An examination of point-to-serial correlations revealed that among physicians in experimental firms, improvement in scores from pretest to posttest was strongly related to seminar attendance ($r = 0.67$, $p < 0.0006$) but was not correlated with the use of preventive procedures. Seminar attendance alone similarly was not correlated with the use of preventive procedures.

Correlations between the use of the three preventive measures were high (Mammography and influenza vaccine $+0.49$ $p = 0.002$; mammography and pneumovax $+0.52$ $p = 0.001$; and pneumovax and influenza $+0.91$ $p < 0.0001$).

Discussion

Much of current medical practice is based upon precedent alone. The application of scientific rigor to the evaluation of the effectiveness and efficacy of various preventive and therapeutic regimens, as well as of systems for their implementation is a relatively recent development. Not surprisingly, in this context evaluative data often has had less clinical impact than may be appropriate. Consequently we frequently observe high utilization of interventions of questionable efficacy, while those of demonstrated value may be consistently underutilized. The causes of this discrepancy between ideal and actual practice are manifold and relate to issues of technology diffusion, knowledge, and attitudes of practitioners and patients, and to a wide variety of incentives determined by the personal, professional, social, and economic environment.

In this study we examined the use of a small number of preventive interventions of well-documented efficacy.[6–8] These procedures and preventive maneuvers have been evaluated carefully in the literature and were found to be safe and effective. Guidelines for their use have been clearly defined and target populations upon whom specific interventions would have the most impact have been determined.

Because the effectiveness of these preventive interventions has been well documented in the literature, we felt that it was safe to assume that the process of immunization and screening would result in desired outcomes. The efficacy of the recommended screening procedures has been demonstrated in detecting diseases at a point early enough to alter their natural histories. Few medical practices have been submitted to such intense scrutiny. It follows that compliance with these guidelines should be part of standard medical practice. Surprisingly, this has not been the case. Confusion exists with respect to what items should be included in the periodic health examination and the frequency with which they should be performed.[2]

In our clinics we found that, with the exception of blood pressure measurements, routine screening and preventive procedures rarely were carried out. We therefore proceeded to design an intervention to improve house staff compliance with preventive medicine guidelines. The intervention itself purposely was kept simple, relying upon a minimum of on-site staff supervision and monitoring. What we observed was a rather remarkable change in the house officers' attitudes and use of preventive procedures. It was interesting to see that the effect of the educational seminars seemed far less important than the simple checklists in changing physician behaviors. It appeared that a change in attitude, rather than in specific factual knowledge, was correlated most closely with the change in use of preventive procedures. Further research will be necessary to determine what factors influence attitude most directly. Similarly, further investigation will be necessary to determine the duration of the effect of our intervention.

It may be, however, that an intervention of the sort we have described should be repeated on an annual basis in teaching hospitals. As with all new information and diffusing technologies, simple knowledge of efficacy and effectiveness does not ensure appropriate use. A constant reminder, perhaps in the form of a checklist, may be necessary to improve the delivery of certain procedures that should be, but often are not, regularly considered by physicians. A system to ensure the periodic application of screening and preventive maneuvers of demonstrated effectiveness clearly can contribute to the provision of quality care.

References

1. Cohen DI, Jones P, Littenberg B, Neuhauser D. Does cost information availability reduce physician test usage? A randomized clinical trial with unexpected findings. Med Care 1982;20:206.

2. Romm FJ, Fletcher SW, Hulka BS. The periodic health examination: Comparison of recommendations and internists' performance. South Med J 1981;74(3):265.

3. Carter WB, Belcher DW, Inui TS. Implementing preventive care in clinical practice: problems for managers, clinicians, and patients. Med Care Review 1981;38(4):195.

4. Waggoner DM, Frengley JD, Griggs RC, Rammelkamp CH. A "firm" system for graduate training in internal medicine. J Med Educ 1979;54:556.

5. Lampman JH. Unpublished data.

6. Spitzer WO. Report of the Task Force on the Periodic Health Examination. Can Med Assoc J 1979;121:1193.

7. American Cancer Society. Guidelines for the cancer related checkup. Recommendations and rationale. CA-A Cancer Journal for Clinicians 1980;30(4):194.

8. Breslow L, Somers AR. The lifetime health-monitoring program. N Engl J Med 1977;296:601.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Communication

Spending Time With Patients

The Impact of Organizational Structure on Medical Practice

FREDRIC D. WOLINSKY, PH.D.,* AND WILLIAM D. MARDER, A.M.†

This article assesses whether the amount of time that physicians spend overall, in the office, and in the hospital per patient visit depends upon the organizational structure of the physician's practice (i.e., does the physician practice in the fee-for-service system or in an HMO, and if in an HMO, in what type of an HMO). Data pooled from two national studies (N = 2,521) reveal several interesting patterns, including: 1) on all measures, solo physicians spend more time per patient visit than physicians in group practice; 2) overall and in the office, internists spend more time per patient visit than general practitioners; 3) overall and in the hospital, obstetricians-gynecologists spend more time per patient visit than general practitioners; 4) overall and in the office, physicians in prepaid group practices spend less time per patient visit than physicians in fee-for-service group practices; 5) overall and in the office, physicians reimbursed on a straight salary basis spend more time per patient visit than nonsalaried physicians; and, 6) in the hospital, physicians in staff model HMOs spend more time per patient visit than physicians in fee-for-service group practice. The implications of these findings for future studies of the effects of organizational structure on medical practice are discussed.

GIVEN THE EMPHASIS on improving access to ambulatory medical care while containing the rapidly rising costs of such care, Health Maintenance Organizations (HMOs) are receiving more currency as the most palatable competitive approach for the 1980s.[1,2] Indeed, the health services research literature is replete with current and extensive assessments of the promise and performance of HMOs.[3–6] One of the most commonly assumed yet seldom documented attributes of HMOs is the fact that HMO enrollees must wait longer to see a physician and then get to spend less time with the physician than do their fee-for-service counterparts.

In one of the rare comparative studies of this alleged attribute, Mechanic found that "fee-for-service physicians spend more time in direct patient care activities than those in prepaid practice, and devote more time to each patient."[7,8] The explanation

---

* Associate Professor, Hospital and Health Care Administration; Director, Doctoral Program in Health Services Research, Saint Louis University Medical Center.

† Research Associate, Center for Health Policy Research, American Medical Association.

The views presented in this paper are those of the authors and do not necessarily represent the official policy or position of either the Saint Louis University Medical Center or the American Medical Association.

Reprint requests: Fredric D. Wolinsky, Ph.D., Saint Louis University Medical Center, 3525 Caroline Street, St. Louis, Missouri 63104.

0025-7079/82/1000/1051/$01.25 © J. B. Lippincott Co.

he gives for this generally accepted finding is that[8]:

> . . . the patient load characteristic of general practice in pre-paid groups encourages a more assembly line practice which is less responsive to patients than the pattern characteristics of fee-for-service practice. Pre-paid physicians work during scheduled hours and may deal with increased load by processing patients more rapidly. Fee-for-service physicians tend to respond to increased demand by working longer hours.

Two tacit assumptions are operative here. On the one hand, it is assumed that physicians in prepaid group practices have no financial incentive to adjust to increased patient demand by lengthening their work week. On the other hand, it is assumed that fee-for-service physicians can adjust to increased patient demand by lengthening their work week because they will receive additional compensation.

Essentially, this argument assumes the existence of what may be called a target-time hypothesis. That is, although physicians are willing, for the most part, to adjust their practice schedules to patient demands, they will do so only within what they consider to be a reasonable range of hours worked per week. Based on this tacit assumption and in the absence of their receiving additional compensation for practicing additional hours beyond the reasonable range, Mechanic posited that: 1) salaried physicians would adjust their practices to the increase in patient demand by reducing the amount of time that they spend per patient, while 2) physicians in fee-for-service practice would respond by lengthening their work week in return for additional compensation. The data that Mechanic presented were consistent with his expectations,[7,8] although seemingly comparable unpublished data from Held and Reinhardt's Group Practice Study indicate that office visits are generally "about the same length in both practice modes."[9]

In a comparative study of the medical practices of physicians in different types of HMOs,‡ Wolinsky and Corry examined variations in the amount of time these physicians spent with patients.[10] Their findings indicate that in staff model HMOs (wherein physicians are likely to be on a straight salary) physicians spent more time per patient in the office and in the hospital than physicians in either group model or independent practice associations (IPA) model HMOs. Specifically their data indicate that the staff model HMO physicians spent from 10 to 20 per cent more time with their patients than did other HMO physicians.

---

‡ Basically, there are three types of HMOs: staff models, group models, and individual practice associations or IPAs. In a group model HMO all of the physicians are members of the same group practice. Their medical group contracts with the HMO to provide its services, and is thus a separate legal and fiscal entity from the HMO. Usually, the group services that HMO exclusively, and either owns or contracts with the hospital(s) to which its patients are admitted. Members of the group are reimbursed on a salary or salary plus profit-sharing basis. The key element, then, in group model HMOs is that the physicians have a clear proprietary interest in the success or failure of the group and of the HMO. In essence, they are partners in its financial success or failure. The classic example of group model HMOs is the Kaiser Foundation Health Plan.

Unlike physicians in group models, physicians in staff model HMOs are employed directly by the HMO itself. As a result they are employees rather than partners. Therefore, they do not have as clear a proprietary interest in the HMO. They are usually reimbursed on a salary basis, although adjustments may be made based on average patient loads. The classic example of the staff model HMO is the Group Health Plan of Puget Sound.

The IPA is a loose federation of independent, individual physicians who agree to treat patients enrolled in a third party's HMO on a fee-for-service basis in their own offices. Frequently, the IPA is jointly sponsored by the local medical societies and insurance companies. When an IPA physician provides services to an HMO enrollee, the physician is reimbursed by the IPA for services rendered, less a modest discount rate for the administration of the plan. Unlike staff and group model HMO physicians, who usually have the majority of their patients coming from the HMO, IPA physicians usually have no more than 10 per cent of their patients coming from the HMO. As a result, while the HMO practice is usually the major if not the entire practice for staff and group model physicians, it is a small sideline for physicians in IPAs. The classic example of an IPA is the Forbes Health Maintenance Plan of Pittsburgh.

In explaining their findings, Wolinsky and Corry posited that the longer patient visits provided by staff model HMO physicians reflect the absence of financial incentives motivating them to provide either more services in general or fewer services per enrollee, while seeing more enrollees. Accordingly, they argue that:

> . . . because staff model HMO physicians are principally reimbursed on a straight salary, there is no financial incentive impinging upon them to reduce the amount of time they spend with each patient. As a result, staff model HMO physicians appear to be more willing to spend significantly longer periods seeing their patients in both the office and hospital settings.[10]

Thus, like Mechanic, Wolinsky and Corry focus on an economic explanation, although their emphasis is on the absence rather than the presence of financial incentives that would motivate staff model HMO physicians to increase the demand for their services.

Clearly the data presented by Mechanic and by Wolinsky and Corry contradict each other and lead to different, albeit economically oriented, interpretations. Indeed, both interpretations appear to be straightforward and plausible, but a theoretic reconciliation is not obvious. Empirically, however, the results of the two studies may vary because of differences in design that lead to alternative analytic shortcomings. For example, Mechanic's study directly compares fee-for-service physicians with physicians in prepaid group practices, assuming that all HMOs are basically alike and modeled after the traditional group model typified by Kaiser Permanente. Wolinsky has argued that this is not the case and that the differences in structural configuration between staff model, group model, and IPA model HMOs have significant impacts on both the macro and micro level performance of the plan and its individual physician's practices.[6] The Wolinsky and Corry study supports this claim by showing significant differences in individual physician's medical practices between HMO types. Their data, however, fail to directly compare the three types of HMO physicians with the traditional fee-for-service physician, invoking instead the tacit assumption that physicians working in IPA model HMOs are sufficient proxies for non-HMO fee-for-service physicians. Thus, the strength of Mechanic's study is that it directly compares fee-for-service physicians with HMO physicians, while the strength of the Wolinsky and Corry study is that it identifies the different effects produced by various HMO organizational structures. To be sure, the strength of one study is the weakness of the other, and vice versa. Accordingly, although these studies address the same general issue, they utilize methods and data that do not permit direct comparisons.

In the present study we incorporate the best of the Mechanic and the Wolinsky and Corry studies to assess the impact of the organizational structure of medical practice on the amount of time physicians spend with their patients. To do so we pool cross-sectional data from two national studies of physicians and their medical practices. The pooled data set contains information on the medical practices of 5,133 physicians working in the conventional fee-for-service system, and 1,642 physicians working in HMOs. To control for variations in case-mix, we limit the analysis to a subset of 2,521 primary care physicians, for whom all of the appropriate information had been obtained. We then use multiple regression techniques to determine the impact of organizational structure on the amount of time that physicians spend with their patients overall, in the office, and in the hospital.

Data and Methods

The data on fee-for-service physicians are taken from the 14th annual Periodic Survey of Physicians (PSP-14M) conducted in 1980 by the Center for Health

Services Research and Development of the AMA. The PSP-14M questionnaire was mailed to a 5 per cent, single-stage, proportionately stratified (on medical specialty categories) random sample of nonfederal, office-based physicians engaged in active medical practice in the United States. The sample was taken directly from the AMA Physician Masterfile's enumeration of every physician, regardless of membership, in the United States. By using a series of four follow-up mailings at approximately monthly intervals, a useable completion rate of 50 per cent was obtained providing data on approximately 5,133 physicians.

An analytic comparison of the medical specialty, census division, age, and sex of responding physicians with the physician population revealed no substantive differences.[11] Although the distributions on these known characteristics are demonstrably similar, the comparability of the distributions of the respondents and the non-respondents on the unknown characteristics (e.g., the amount of time spent with patients) remains unknown. However, Barton and her colleagues[12,13] recently have shown that when a reasonable response rate is obtained from a homogeneous population like ours, the estimation of statistical associations such as those reported herein are "unlikely to be affected by major bias due to non-response." Moreover, the most frequent bias identified in the nonresponse literature is that the respondents generally are better educated than the nonrespondents.[14–16] Obviously, this potential bias is not applicable to our study of physicians. Finally, in a more clinical vein, Criqui and his colleagues have shown that when respondents and nonrespondents with similar distributions on exposure variables (e.g., demographic and occupational factors) are examined, comparable prevalence and relative risk estimates are obtained.[17,18] Accordingly, we assume that given our purpose, the high response rate, the homogeneity of the population, and the marked similarity between the known characteristics of the respondents and the nonrespondents, these data are not likely to be unduly distorted by nonresponse biases. Thus, the PSP-14M data provide an appropriate data source with which to empirically assess the impact of organizational structure on the amount of time that physicians spend with their patients in the traditional fee-for-service setting.

Comparable data on physicians practicing in HMOs was obtained in two stages. First, a letter was sent to all HMOs listed in the 1978 census of HMOs compiled by the federal Office of Health Maintenance Organizations (OHMOs).[19] The executive directors of each of the 203 HMOs enumerated in the census were asked to provide a list of all physicians participating in their plans. Ninety-seven of the HMOs responded and an additional 17 lists were obtained from the Register of Group Practice concurrently being assembled by the AMA. An analytic comparison of the response rates for the three major types of HMOs failed to indicate any nonresponse biases. Second, a 20 per cent sample then was selected by systematically choosing one out of every five physicians from each HMO physician list using a new random seed within each HMO. The HMO physician sample then was sent the standard PSP-14M questionnaire during 1980. Using four follow-up mailings to subsequently nonresponding physicians at approximately monthly intervals, a useable response rate of 50 per cent was obtained, resulting in a sample of 1,642 HMO physicians. Based on the most recent categorizations published by the OHMO in the *Federal Register*, responding physicians were classified as participating in group model (N = 294), staff model (N = 163), or IPA model (N = 1,160) HMOs. On all known criteria, which are admittedly few and not specific, the resulting sample of HMO physicians (PSP-14H) appears to be representative of the population of HMO physicians.[10] Thus, PSP-14H represents an appropriate data source with which to assess the impact of organizational structure on the amount of time that physicians in HMOs spend with their patients.

When taken together, the PSP-14M and PSP-14H physician surveys comprise the data set to be used in this paper (hereafter referred to as PSP-14). The pooled PSP-14 data set allows us to compare the impact of organizational structure on the amount of time that physicians spend with their patients in the fee-for-service setting directly with the amount of time that physicians in IPA model, group model, and staff model HMOs spend with their patients. Thus, these data permit us to make the direct comparisons between fee-for-service system physicians and HMO physicians that Mechanic made, as well as distinguishing between the different types of HMO settings examined by Wolinsky and Corry.

In this study the dependent variables are the numbers of minutes spent per patient visit overall, in the office, and in the hospital. These variables were calculated from information contained on the PSP-14 questionnaires concerning the total number of hours worked, the total number of patients seen, the total number of hours worked in the office, the total number of patients seen in the office, the total number of hours worked in the hospital, and the total number of patients seen in the hospital. In all cases, physicians were asked to indicate the total number of hours (or visits) in their most recent, complete week of practice.

Unfortunately, such retrospective self-reporting has a serious limitation and we would be remiss not to discuss it in some detail. This limitation recently has been addressed by Nelson et al.[20] with special reference to physicians. Using techniques logically consistent with classic multitrait-multimethod and triangulation approaches,[21-25] Nelson et al. compared the retrospective self-reports of 13 primary care physicians with checklists obtained from independent shadow-observers over a 1-week period. They failed to find close agreement between the two data sources concerning the frequency and duration of various tasks (e.g., the number and length of patient visits); physicians reported doing more tasks and doing them longer than the observers' records would indicate. As a result, the point estimates (i.e., the means) obtained by using the physicians' retrospective self-reports probably were overstated. This overstatement may have resulted from the physicians including in their actual patient contact estimates the amount of time that they spent reviewing charts, writing case histories, and attending grand rounds. In any event, Nelson et al. argue that the retrospective self-reporting method, at least among physicians, may not be valid for generating point estimates.

While the validity of the point estimates themselves may be questioned (e.g., the amount of time spent with patients may be overestimated), there is no empirical evidence to suggest that this tendency to overestimate is more or less pronounced across the different organizational settings in which physicians may practice. That is, physicians in staff model HMOs are no more or no less likely to over- or underestimate the amount of time that they spend with their patients than are solo practitioners or physicians in fee-for-service group practices. Thus, the differences between the point estimates for various groupings of physicians based on the organizational structure of their medical practices should be reliable and meaningful. Therefore, our use of these data for focusing on the differences in the amounts of time that physicians spend with their patients is indeed appropriate. Accordingly, we focus and report solely on the differences between the point estimates for physicians in the various organizational structures and not on the point estimates themselves.

The principal independent variables were dummies reflecting the organizational setting in where the physician practiced medicine, including: solo fee-for-service, group model HMOs, staff model HMOs, and IPA model HMOs. The dummy variable for the fee-for-service group practice category was omitted to preserve the singularity of the correlation matrix. Thus, the regression coefficient for each of the dummy variables indicates the difference between being in that particular organizational structure as opposed to being in fee-for-service group practice.

In addition to the principal independent variables measuring organizational structure, several others were included to control for a variety of effects. For example, in order to control for case-mix variations (even though we restricted the analysis to primary care physicians), we include dummy variables for internal medicine, pediatrics, and obstetrics-gynecology. The omitted category is that of general and family practice. Thus, the regression coefficients for each of these dummy variables indicates the difference between that particular medical specialty and general or family practice. Similarly, to control for variations in reimbursement methods we constructed a dummy variable that indicated whether or not the physician was paid for the majority of his or her practice on a straight salary basis.§ This partitions the effect of an individual physician's reimbursement from the more macro-level structural effects of the organizational reimbursement setting of his or her medical practice.

Finally, we also controlled for several other factors, including exposure to the HMO (not all HMO physicians work full time for the HMO), the physician's age (as a proxy for life-cycle stage), and the interaction of individual salary reimbursement and working in a staff model HMO (because the combination of a salaried organizational setting and being salaried oneself would be especially reinforcing). None of these exposure, age, and salary-staff model interaction terms, however, had any significant impact on the amount of time that physicians spend with their patients. We mention having controlled for these factors here only for the benefit of the reader.

Findings

Table 1 contains the regression coefficients, intercepts, and squared multiple correlation coefficients (r^2) obtained from the models predicting the amount of time physicians spend with their patients overall, in the office, and in the hospital. Coefficients significantly different from zero at the 0.05 level or beyond are indicated with an asterisk. In general, the model explained significant proportions of the variance in the amount and time physicians spend with their patients, although the predictive power was greater for the overall and office-visit indicators. The particular effects of the specialty, solo fee-for-service, HMO, and reimbursement characteristics are described below.

Specialty Characteristics

As indicated earlier, the regression coefficients for each of the three specialty dummies indicate the difference between that specialty and the control category of general or family practitioners. The significance of the specialty dummies indicates the importance of controlling for case-mix variation when assessing the amount of time physicians spend with their patients. That is, the data indicate that internists spend nearly 11 minutes more, overall, and nearly 10 minutes more in the

---

§ Although it would seem that this reimbursement method variable might create a collinearity problem when used in conjunction with the staff model HMO dummy, this clearly was not the case. The zero-order correlation between these two variables was only 0.203. This low correlation probably reflects the growing tendency of partners to incorporate and then arrange to have themselves placed on a salary by the corporation, thus minimizing their income tax liabilities. Because we simultaneously control for the effects of fee-for-service group practice (our reference category) as well as for the effects of staff model HMOs, this anomaly should neither cloud nor constrain a straightforward interpretation of the salaried reimbursement variable. That is, the fee-for-service group practice reference category should measure the effects of being in fee-for-service group practice (including being salaried in a fee-for-service group practice), and the staff model HMO dummy should measure the effects of being in a salaried reimbursement environment. Thus, the micro-level reimbursement variable should be measuring solely the effects of the individual physician's being salaried.

TABLE 1. Regression Coefficients Obtained From Models Predicting the Amount of Time That 2,521 Primary Care Physicians Spend With Their Patients Overall, in the Office, and in the Hospital, Per Visit.*

Independent Variables	Dependent Variables†		
	Overall	In the Office	In the Hospital
Intercept	23.30	20.11	37.90
Internal medicine	10.85*	9.68*	−3.65
Pediatrics	−0.80	−0.88	5.78
Obstetrics-gynecology	6.07*	0.95	18.70*
Solo F-F-S	3.57*	1.99*	11.47*
IPA model HMO	−0.49	−0.29	−1.44
Group model HMO	−3.19*	−3.58*	2.25
Staff model HMO	1.82	−0.81	14.67*
Salaried reimbursement	3.94*	2.99*	6.28
R^2	0.11*	0.13*	0.04*

* Coefficients significant at the 0.05 level or beyond.

† Measured in minutes per visit. Alternative transformations (including normal logarithms and polynomial expansions) failed to alter the findings, suggesting that minutes per visit is the appropriate measurement unit.

office with each of their patients, than general or family practitioners. That the coefficient for internal medicine on the hospital visits measure is not significant and has a negative sign probably reflects the fact that internists provide more extensive diagnostic services in their offices compared to more brief consultative services in hospitals. Pediatricians, on the other hand, do not spend significantly more or less time than general or family practitioners with their patients in any of the three settings. Finally, obstetricians-gynecologists spend 6 minutes more, overall, and nearly 19 minutes more in the hospital, than their general or family practice counterparts. This probably reflects the greater hospital or surgical orientation of obstetricians-gynecologists when compared to general or family practitioners.

Solo Fee-for-Service

Perhaps one of the more interesting findings portrayed in Table 1 is the significant and positive coefficients that the solo fee-for-service dummy produces on all three dependent variables, confirming Mechanic's earlier work.[8] When compared to their group fee-for-service counterparts, solo fee-for-service physicians spend nearly 4 minutes more with their patients overall, nearly 2 minutes more with their patients in the office, and over 11 minutes more with their patients in the hospital. According to Freidson, the consistently greater amounts of time spent by solo fee-for-service physicians with their patients may well reflect their greater sense of client dependency, as opposed to the greater colleague dependency felt by their group fee-for-service counterparts.[26] That is, Freidson argues that solo practitioners are more directly dependent on their patients for their patient flow (especially when in general or family practice) than their colleagues who band together in some form of collaborative practice arrangements. In essence, physicians in collaborative practice arrangements may receive patients through the secondary benefit of the brand-name effect of the group or on direct referral from their partners; physicians in solo practice do not enjoy the advantages of such secondary benefits or direct referral networks.

The HMO Characteristics

Three rather interesting patterns are reflected in the regression coefficients produced by the HMO organizational types.

First, the regression coefficients indicate that IPA model HMO physicians do not spend significantly more or less time with their patients in any setting than physicians in fee-for-service group practices. This is to be expected, since the IPA model HMO essentially is a sideline for its physicians and usually amounts to 10 per cent or less of their total medical practices. Therefore, although legally associated with an HMO, IPA physicians are not likely to be affected by its incentive structures. Second, the coefficients for group model HMO physicians provide some support for Mechanic's interpretation regarding the total and office visits measures. Physicians in group model HMOs spend significantly less time (about 3–4 minutes) than their counterparts in fee-for-service group practices. Thus, as Mechanic suggested, group model HMO physicians appear to adjust to increased patient demand by shortening the amount of time they spend with each patient, rather than increasing their work week as do their fee-for-service group practice counterparts. Finally, however, this trend is contradicted by the significantly greater amount of time (approximately 15 minutes) that staff model HMO physicians spend with their patients in the hospital. These data indicate that physicians in staff model HMOs, who generally are salaried employees, spend more time with their patients (at least in the hospital) than their fee-for-service group practice counterparts, who have a proprietary interest in generating practice incomes.

Reimbursement Characteristics

The regression coefficients for the salaried reimbursement method indicate that salaried physicians spend nearly 5 minutes more with their patients overall, and nearly 3 minutes more with their patients in the office than their nonsalaried counterparts. Like the effects of being in a staff model HMO, the effects of being reimbursed on a salaried basis support Wolinsky and Corry's argument and contradict Mechanic's hypothesis. Mechanic posited that salaried physicians would adjust to increased patient demand by spending less time with their patients, which clearly is not the case. Wolinsky and Corry, however, argued that in the absence of fiscal incentives encouraging physicians to see more different patients more often, salaried physicians would spend more time with their patients.

Summary

Unfortunately the findings reported above are not sufficient to resolve the controversy surrounding the amount of time that physicians in different organizational settings spend with their patients. While some of the findings support Mechanic's interpretation (i.e., that group model HMO physicians generally spend less time with their patients than physicians in fee-for-service group practices), others support Wolinsky and Corry's interpretation (i.e., salaried physicians spend more time with their patients than nonsalaried physicians). These data do, however, provide several valuable insights concerning the relationship between organizational structure and individual physicians' practices. First, the difference in the coefficients for the staff model HMO dummy and the salaried reimbursement variable indicate that there is, indeed, a difference between being salaried and working in a predominantly salaried environment. Moreover, these data indicate that the micro-level incentive (i.e., that the physician is on a salary) is more effective in altering individual physicians' medical practices than the macro-level incentive (i.e., merely working in an environment where most of one's colleagues are salaried). Second, it is important to control for the effects of case-mix when examining the impact of organizational structure on individual physicians' practices, as indicated by the substantial regression coefficients for the medical specialty dummies. Finally, the medical practices of physicians working in IPA model HMOs are not significantly different from those of their counterparts in fee-for-service group practices. Therefore, the use of data on physicians in IPAs as a proxy for data on fee-for-service physicians when making comparisons to data on physicians in staff and group model HMOs may well be appropriate.

Implications

Two implications for future research and policy may be drawn from the findings reported above. First, future research must explicitly identify and measure the specific incentives that, when bundled together, constitute the various organizational structures of medical practice. Until these specific incentives have been identified and measured, it will not be possible to determine exactly why medical practice varies from one organizational structure to the next.[6] Second, future research also must be based upon the analysis of longitudinal rather than cross-sectional data in order to determine whether there is a selection process that brings fundamentally different types of physicians to practice in the different organizational structures.[26] Although the cross-sectional data used herein, as well as that used by Mechanic, and by Wolinsky and Corry, have been useful in establishing the relationship between variations in organizational structures and differences in individual physicians' medical practices, clear causal inferences only can be made from longitudinal data.

References

1. Enthoven A. Consumer choice health plan. N Engl J Med 1978;298:650.

2. PL 93-641. The national health planning and resources development act of 1974. Washington, D.C.: U.S. Government Printing Office. 1974.

3. Blumberg M. Health status and health care use by type of private health coverage. Milbank Mem Fund Quarterly 1980;58(4):633.

4. Iglehart J. The federal government as venture capitalist: how does it fare? Milbank Mem Fund Q 1980;58(4):656.

5. Luft H. Assessing the evidence on HMO performance. Milbank Mem Fund Q 1980;58(4):501.

6. Wolinsky F. The performance of HMOs: an analytic review. Milbank Mem Fund Q 1980;58(4):537.

7. Mechanic D. General medical practice: some comparisons between the work of primary care physicians in the United States and England and Wales. Med Care 1972;10:402.

8. Mechanic D. The organization of medical practice and practice orientations among physicians in prepaid and nonprepaid primary care settings. Med Care 1975;13:189.

9. Held P, Reinhardt U. Analysis of economic performance in medical group practices. Princeton: Mathematica Policy Research, 1979.

10. Wolinsky F, Corry B. Organizational structure and medical practice in health maintenance organizations. In: Goldfarb D, ed. Profile of medical practice, 1981. Chicago: American Medical Association, 1981.

11. Goldfarb D, ed. Profile of medical practice, 1981. Chicago: American Medical Association, 1981.

12. Barton J, et al. Characteristics of respondents and non-respondents to a mailed questionnaire. Am J Public Health 1980;70(8):823.

13. Belanger C, et al. The nurses' health study. Am J Nurs 1978;78(9):1038.

14. Suchman E, McCandless B. Who answers questionnaires? J Appl Psychol 1940;24(7):758.

15. Edgerton H, et al. Objective differences among various types of respondents to a mailed questionnaire. Am Sociol Rev 1947;12(4):435.

16. Reuss C. Differences between persons responding and not responding to a mailed questionnaire. Am Sociol Rev 1943;8(4):433.

17. Criqui M, et al. Differences between respondents and non-respondents in a population-based cardiovascular disease study. Am J Epidemiol 1978;108(4):367.

18. Criqui M. Response bias and risk ratios in epidemiologic studies. Am J Epidemiol 1979;109(4):394.

19. U.S. Department of Health, Education, and Welfare. Census of prepaid plans, 1978. Washington, D.C.: U.S. Government Printing Office, 1978.

20. Nelson E, et al. A study of the validity of the task inventory method of job analysis. Med Care 1975;13(2):104.

21. Campbell D, Fiske D. Convergent and discriminant validation by the multitrait, multimethod matrix. Psychol Bull 1969;56(1):81.

22. Denzin N. The research act. 2nd ed. New York: McGraw-Hill, 1978.

23. Campbell D. Social attitudes and other acquired behavioral dispositions. In: Koch S, ed. Psychology: a study of science, vol 6. New York: McGraw-Hill, 1963.

24. Webb E. Unconventionality, triangulation and inference. In: Proceedings of the 1966 invitational conference on testing problems. Princeton: Educational Testing Service, 1966.

25. Webb E, et al. Unobtrusive measures: nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

26. Freidson E. The profession of medicine. New York: Dodd, Mead, 1970.